ABSTRACT

KRACHEY, ELIZABETH C. Variations on the Accelerated Failure Time Model: Mixture Distributions, Cure Rates, and Different Censoring Scenarios. (Under the direction of Professors Wenbin Lu and Sujit K. Ghosh).

The accelerated failure time (AFT) model is a popular model for time-to-event data. It provides a useful alternative when the proportional hazards assumption is in question and it provides an intuitive linear regression interpretation where the logarithm of the survival time is regressed on the covariates. We have explored several deviations from the standard AFT model.

Standard survival analysis assumes that in the case of perfect follow-up, every patient will eventually experience the event of interest. However, in some clinical trials, a number of patients may never experience such an event, and in essence, are considered cured from the disease. In such a scenario, the Kaplan-Meier (KM) survival curve will level off at a nonzero proportion. Hence there is a window of time in which most or all of the events occur, while heavy censoring occurs in the tail. The two-component mixture cure model provides a means of adjusting the AFT model to account for this cured fraction. Chapters 1 and 2 propose parametric and semiparametric estimation procedures for this cure rate AFT (CRAFT) model. Parametric estimation is employed to verify the performance of the CRAFT model in a simpler setting. Semiparametric estimation using mixture density estimation is then used to introduce a novel estimation procedure for the CRAFT model. Posterior consistency is established under some regularity conditions providing large sample justification to the proposed model. Simulation studies validate the performance of the proposed model in finite samples and an application to a breast cancer data set illustrates the semiparametric estimation procedure.

Survival analysis methods for interval-censoring have been much slower to develop than for the right-censoring case. This is in part because interval-censored data have a more complex censoring mechanism and because the counting process theory developed for right-censored data does not generalize or extend to interval-censored data. Because of the analytical difficulty associated with interval-censored data, recent estimation strategies have focused on the implementation rather than the large sample theoretical justifications of the semiparametric AFT model. Chapter 3 proposes a semiparametric Bayesian estimation
procedure for the AFT model under interval-censored data, employing Markov chain Monte Carlo methods to generate samples from the posterior distribution. Posterior consistency is established under certain conditions providing large sample justification to the proposed model. Simulation studies validate the performance of the proposed model in finite samples and an application to a breast cosmesis data set illustrates the method.
Variations on the Accelerated Failure Time Model: Mixture Distributions, Cure Rates, and Different Censoring Scenarios

by
Elizabeth C. Krachey

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics
Raleigh, North Carolina
2009

APPROVED BY:

Dr. Marie Davidian
Dr. Brian Reich

Dr. Wenbin Lu
Chair of Advisory Committee
Dr. Sujit K. Ghosh
Co-Chair of Advisory Committee
DEDICATION

To my husband, Matthew
and to my parents, Dan and Cathy
Elizabeth (Liz) Catherine Krachey was born to Daniel and Catherine Nelson on May 16, 1982. After a childhood amidst the evergreen trees, lakes and waterways, mountain views, and rain and shine of Seattle, Liz moved cross-state to the rolling hills, wheat fields, and star-filled skies of Walla Walla. There she played varsity collegiate volleyball while receiving her Bachelor of Arts degree in mathematics from Whitman College. Next, Liz decided to move cross-country to experience the South. Amidst country line dancing, crazy humidity, bountiful farmers markets, and the waterfalls and bluegrass of Appalachia, Liz received her Master of Science degree in statistics from North Carolina State University. And this dissertation completes the requirements for her Doctor of Philosophy degree in statistics, also at North Carolina State University. Liz is the oldest of six and is married to Matthew James Krachey.
ACKNOWLEDGMENTS

Thank you to my advisers– Wenbin Lu and Sujit K. Ghosh, for your encouragement, guidance, generosity, and many good ideas. I feel incredibly grateful to have been surrounded by your never-ending supplies of optimism throughout the dissertation process. Also thank you to my other committee members– Marie Davidian and Brian Reich, for your willingness to help and share ideas along the way.

Thanks to the statistics department at North Carolina State University, for taking me on and providing funding opportunities. Also to Jeff Sorbel and Chuck Wakeford at Gilead Sciences, for giving me a great introduction to pharmaceuticals while also being extremely encouraging and understanding of my dissertation work. Thank you Adrian Raftery for suggesting North Carolina State University– without your recommendations I would not be here. And a special thank you to Terry Byron and Chris Waddell. No matter how many technical questions or problems I came up with, you each always gave me a friendly face and a relentless persistence in finding a solution.

Also thank you Assumption St. Bridget School, Holy Names Academy, and Whitman College; you each have fostered an excellent well-rounded education emphasizing that learning (intellectually, morally, socially, spiritually) takes place both in and out of the classroom. I hope this lesson stays with me for the rest of my life.

I feel very grateful for the close friendships that I have made while in graduate school– who have kept me going and really helped me tough it out for the long haul. Especially thank you Laine Thomas, Steven Thomas, Kelci Miclaus, Funda Gunes, Casey Theriot, and Jason Brinkley. Also thank you Marion Rubali, Elly Smith, Molly Mayes, and Sarah Sitts– for keeping in touch and reminding me that I am loved, especially when I first set out to a whole new part of the country. Thank you to my family, Mom, Dad, Joe, Kelly, Mark, Zoe, and Jack, for your unconditional love and support. Our weekly Sunday calls have really helped me keep going while also keeping me up to par with the family gossip. And lastly, Matthew, I am forever grateful for your love and belief in me.
# TABLE OF CONTENTS

LIST OF TABLES ................................................................. vii

LIST OF FIGURES ............................................................... ix

1 The Parametric CRAFT model Under Right Censoring ........................ 1
   1.1 Introduction ............................................................ 1
   1.2 The Parametric CRAFT Model ......................................... 3
   1.3 Observed Likelihood .................................................. 4
   1.4 Estimation .............................................................. 4
      1.4.1 Maximum Likelihood ............................................ 4
      1.4.2 EM Algorithm ................................................... 5
      1.4.3 Bayesian Methodology ......................................... 6
   1.5 Simulation Studies ................................................... 7
   1.6 Conclusion ............................................................ 9

2 The Semiparametric CRAFT model Under Right Censoring ................ 11
   2.1 Introduction ............................................................ 11
   2.2 The CRAFT Model with Unspecified Error Distribution ............... 12
   2.3 Implementation ........................................................ 13
   2.4 Semiparametric Estimation Using the Posterior Distribution ........ 15
   2.5 Consistency of Posterior Distribution ................................ 17
   2.6 Simulation Studies ................................................... 19
   2.7 Breast Cancer Data Set ............................................... 21
   2.8 Conclusion ............................................................ 26

3 The Semiparametric AFT Model Under Interval Censoring .................. 27
   3.1 Introduction ............................................................ 27
   3.2 The AFT Model with Unspecified Error Distribution ................ 28
   3.3 Prior Distribution .................................................... 30
   3.4 Semiparametric Estimation Using the Posterior Distribution ........ 30
   3.5 Consistency of Posterior Distribution for Case 1 Interval Censoring 32
   3.6 Simulation Studies ................................................... 34
   3.7 Breast Cosmesis Data Set ............................................ 36
   3.8 Conclusion ............................................................ 39

Bibliography ................................................................. 40

Appendices ................................................................. 48
Appendix A Supplement to Chapter 1. ................................................. 49
  A.1 Derivation of Standard Errors based on the EM algorithm ............... 49
  A.2 Densities and Derivatives of the Error Distribution .................... 51
  A.3 Simulation Results for Frequentist Estimation Methods ................. 52
  A.4 Simulation results for Bayesian Estimation Methods .................. 56

Appendix B Supplement to Chapter 2. ................................................. 59
  B.1 Explanation of the Difference Operator Matrix ......................... 59
  B.2 Proof of Posterior Consistency ........................................... 60
  B.3 WinBUGS Code for the Semiparametric CRAFT Model Under Right Censoring 64
  B.4 Simulation Results ......................................................... 66
  B.5 Figures from the Breast Cancer Data Set ................................ 69

Appendix C Supplement to Chapter 3. ................................................. 71
  C.1 Proof of Posterior Consistency ........................................... 71
  C.2 WinBUGS Code for the Semiparametric AFT Model Under Interval Censoring 78
  C.3 Simulation Results ......................................................... 80
  C.4 Figures from the Breast Cosmesis Data Set ............................. 82
Table 2.1 Breast cancer data: Bayesian posterior summaries for both the latency and incidence statistics associated with treatment A or B, clinical stage indication, and number of lymph nodes in the breast cancer data set using the semiparametric CRAFT model ................................................................. 23

Table 3.1 Bayesian posterior summaries for beta (treatment parameter) in the breast cosmesis data set using the semiparametric AFT model ......................... 37

Table A.1 Densities and their respective first and second derivatives for the error distribution .............................................................. 51

Table A.2 Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the standard normal distribution ............................................. 52

Table A.3 Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the extreme value distribution ................................................. 53

Table A.4 Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the logistic\((0,1)\) distribution .............................................................. 54

Table A.5 Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the logistic\((0,0.7)\) distribution .............................................................. 55

Table A.6 Simulation results for the parametric CRAFT model based on Bayesian estimation with 500 Monte Carlo simulations when \( n = 100 \), \( \gamma_0 = 1.0 \) in the data generation, and we set the \( \gamma_0 \) prior to be \( \gamma_0 \sim N(0,10) \) ................................................. 56
Table A.7 Simulation results for the parametric CRAFT model based on Bayesian estimation with 500 Monte Carlo simulations when \( n = 100 \) or \( 200 \), \( \epsilon \sim N(0,1) \) or logistic\((0,1)\), \( \gamma_0 = 1.0 \) in the data generation, and we set the \( \gamma_0 \) prior to be \( \gamma_0 \sim N(0,10) \). ................................................................. 57

Table A.8 Simulation results for the parametric CRAFT model based on Bayesian estimation with 500 Monte Carlo simulations when \( n = 100 \) or \( 200 \), \( \epsilon \sim N(0,1) \) or logistic\((0,1)\), \( \gamma_0 = 0.5 \) in the data generation, and we set the \( \gamma_0 \) prior to be \( \gamma_0 \sim N(0,10) \). ................................................................. 58

Table B.1 Simulation results for the semiparametric CRAFT model based on Bayesian estimation with 1000 Monte Carlo simulations when \( n = 100 \) or \( 200 \), \( \gamma_0 = 0.5 \) or \( 1.0 \) in the data generation, and \( \epsilon \) is generated from a symmetric logistic distribution... 66

Table B.2 Simulation results for the semiparametric CRAFT model based on Bayesian estimation with 1000 Monte Carlo simulations when \( n = 100 \) or \( 200 \), \( \gamma_0 = 0.5 \) or \( 1.0 \) in the data generation, and \( \epsilon \) is generated from a skewed extreme value distribution. 67

Table B.3 Simulation results for the semiparametric CRAFT model based on Bayesian estimation with 1000 Monte Carlo simulations when \( n = 100 \) or \( 200 \), \( \gamma_0 = 0.5 \) or \( 1.0 \) in the data generation, and \( \epsilon \) is generated from a bimodal distribution......... 68

Table C.1 Simulation results for the semiparametric AFT model under case 1 interval censoring with 100 Monte Carlo simulations. Sample sizes are \( n = 200 \) or \( 400 \), and \( \epsilon \) is generated from normal, logistic, extreme value or bimodal distributions. ...... 80

Table C.2 Simulation results for the semiparametric AFT model under case 2 interval censoring with 100 Monte Carlo simulations. Sample sizes are \( n = 200 \) or \( 400 \), and \( \epsilon \) is generated from normal, logistic, extreme value or bimodal distributions. ...... 81
LIST OF FIGURES

Figure 2.1 Estimated (based on semiparametric CRAFT model) and true survival probabilities for data generated from logistic, extreme value, and bimodal distributions when \( n = 200 \) and \( \gamma_0 = 1 \) for Monte Carlo simulations of size 1000. .......................... 20

Figure 2.2 Breast cancer data set: Kaplan–Meier survival and the logarithm of the cumulative hazard function curves for each of the three treatments. ....................... 22

Figure 2.3 Breast cancer data: survival curves using (i) posterior median of the survival estimates based on the semiparametric CRAFT model (weighted by lymph nodes and clinical stage) and (ii) nonparametric Kaplan-Meier estimator, for each of the three treatments. ................................................................. 24

Figure 2.4 Breast cancer data: posterior median of the error density estimates from the semiparametric CRAFT model using a mixture of 17 normal distributions ....... 25

Figure 3.1 Estimated (based on semiparametric AFT model) and true survival probabilities for data generated from either a logistic, extreme value, or bimodal error distribution, with case 2 interval censoring and \( n = 400 \) for Monte Carlo simulations of size 100. .......................................................... 36

Figure 3.2 Interval-censored breast cosmesis data: survival curves using (i) posterior median of the survival estimates based on the semiparametric AFT model and (ii) Turnbull nonparametric estimator, for each of the two treatments. ............... 38

Figure 3.3 Interval-censored breast cosmesis data: posterior median error density estimates from the semiparametric AFT model using a mixture of 17 logistic distributions................................................................. 38

Figure B.1 Breast cancer data: posterior means of the mixing weights using the semiparametric CRAFT model .......................................................... 69

Figure B.2 Breast cancer data: trace and posterior density curves for both the latency and incidence statistics associated with treatment A or B, clinical stage indication, and number of lymph nodes using the semiparametric CRAFT model ............. 70

Figure C.1 Interval-censored breast cosmesis data: trace and posterior density curve for beta (treatment parameter) using the semiparametric AFT model ............ 82
Figure C.2 Interval-censored breast cosmesis data: posterior means of the mixing weights using the semiparametric AFT model ........................................ 83
Chapter 1

The Parametric CRAFT model
Under Right Censoring

1.1 Introduction

Standard survival analysis assumes that in the case of perfect follow-up, every patient will eventually experience the event of interest. However, in some clinical trials, a number of patients may never experience such an event, and in essence, are considered cured from the disease. In such a scenario, the Kaplan-Meier (KM) survival curve will level off at a nonzero proportion. Hence there is a window of time in which most or all of the events occur, while heavy censoring occurs in the tail. Some recent applications of the cure model include Withers et al. (1995), Burnett et al. (2000), Secasan et al. (2005), and Dickman and Adami (2006).

Cure models can either account for short-term and long-term effects together through use of a promotion time cure model (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen et al., 1999; Zeng et al., 2006) or separately model these two pieces through a two-component mixture cure model (Berkson and Gage, 1952). We focus on the mixture cure model which consists of two components, the incidence and the latency. Incidence refers to whether the event will eventually occur and latency indicates lifetime of the event given the subject is susceptible to the event. The incidence rate is commonly modeled by logistic regression. However, many other link functions (e.g., probit, complementary log-log, etc) can also be used to measure the effect of predictions on incidence rates. Several methods have
been explored to model the latency. Farewell (1982) originally suggested parametrically modeling the latency by a Weibull distribution. Taylor (1995) uses a Kaplan-Meier type approach for the latency component, while Kuk and Chen (1992), Sy and Taylor (2000), and Peng and Dear (2000) use Cox’s proportional hazards model (Cox, 1972). However, whether due to time-dependent covariates, crossing KM survival curves, or other reasons, in some cases model inspection suggests that the proportional hazards assumption is possibly incorrect (Patel et al., 2006; Morel et al., 2001), and other ways of modeling the latency are then needed. Yamaguchi (1992), Li and Taylor (2002), and Zhang and Peng (2007) explore semiparametric accelerated failure time (AFT) models, as the AFT model is a popular model when the proportional hazards assumption is in question. Additionally, the AFT model has an intuitive linear regression interpretation where the logarithm of the survival time is regressed on the covariates. Wei (1992) provides an extensive review of the benefits and uses of the AFT model.

Fewer cure models have been explored from a Bayesian framework. Using the promotion time cure model, Chen et al. (1999), Ibrahim et al. (2001), and Yakovlev and Tsodikov (1996) modeled the entire population with a proportional hazards model, and accounted for covariates through a Poisson regression and Weibull density. Kim et al. (2007) extend these results by constructing piecewise-constant exponential terms for modeling the latency. None of these papers provide large sample theoretical justification. Several authors have investigated the AFT model for the general right-censored data (without a cure fraction), including Christensen and Johnson (1988), Johnson and Christensen (1989), Kuo and Mallick (1997), Walker and Mallick (1999), Campolieti (2001), and Hansen and Johnson (2004). Ghosh and Ghosal (2006) established the posterior consistency based on a class of mixture of Weibull densities. However, to the best of our knowledge, in a Bayesian context posterior consistency has not been established for an AFT model with a cure fraction.

The remainder of this chapter will introduce the cute-rate AFT (CRAFT) model and explore parametric estimation procedures. We will verify by simulation studies that the CRAFT model is performing well in these cases. Upon validation, Chapter 2 will propose a semiparametric estimation method for the CRAFT model.
1.2 The Parametric CRAFT Model

We use a logistic model for the incidence and a parametric accelerated failure time model for the latency. Let $\eta$ indicate whether a subject is susceptible ($\eta = 1$) or not ($\eta = 0$) to the event of interest. Let $T$ be the time to occurrence of the event which can be represented as $T = T^* \eta + \infty \cdot (1 - \eta)$ where $T^*$ denotes the latent survival time when subjects are susceptible and we interpret $0 \cdot \infty = 0$. Then the incidence rate can be modeled as

$$P(\eta = 1 \mid z) = p(\tilde{\gamma}, z) = F_0(\tilde{\gamma}^T z),$$

(1.1)

where $\tilde{\gamma} = (\gamma_0, \gamma^T)$ is a $(p+1)$-dimensional vector of unknown parameters, $\tilde{z} = (1, z^T)^T$, $z$ is a $p$-dimensional vector of covariates, and $F_0(\cdot)$ is a known link function, usually chosen to be a cumulative distribution function. We use the logistic link function where $F_0(u) = \{1 + \exp(-u)\}^{-1}$. Notice that many other link functions such as the probit or complementary log-log could have been used in place of the above logistic link or $p(\tilde{\gamma}, z)$ could also have been modeled nonparametrically. Next, the distribution of the latent survival time $T^*$ can be expressed by the following AFT model:

$$\log T^* = \tilde{\beta}^T \tilde{z} + \epsilon,$$

(1.2)

where $\tilde{\beta} = (\beta_0, \beta^T)$ is a $(p+1)$-dimensional vector of unknown incidence parameters and $\epsilon$ represents measurement error. We assume that the same vector of covariates $z$ is present in the incidence and latency components. For computational ease and to verify that the model provides reliable estimates in a simple setting, we proceed with parametric assumptions on the error term. Hence, let $\epsilon \sim S_\epsilon(\cdot)$ where $S_\epsilon(\cdot)$ is a parametrically specified survival function of $\epsilon$. Having defined both the incidence and latency, the CRAFT model is

$$P(T > t \mid Z = z) \equiv S(t \mid \tilde{\beta}, \tilde{\gamma}, z) = 1 - p(\tilde{\gamma}, z) + p(\tilde{\gamma}, z)P(T^* > t \mid Z = z),$$

(1.3)

where $P(T^* > t \mid Z = z) = S_\epsilon(\log t - \tilde{\beta}^T z)$ is the parametric survival function for $T^*$. Hence, the corresponding density and survival functions for survival times $T = t$ is

$$f(t \mid \tilde{\beta}, \gamma, z) = p(\tilde{\gamma}, z)f_\epsilon(\log t - \tilde{\beta}^T \tilde{z})/t,$$

(1.4)

$$S(t \mid \tilde{\beta}, \tilde{\gamma}, z) = 1 - p(\tilde{\gamma}, z) + p(\tilde{\gamma}, z)S_\epsilon(\log t - \tilde{\beta}^T \tilde{z}).$$

(1.5)

Notice that Equation (1.5) is identifiable for any given $S_\epsilon(\cdot)$ provided $p(\tilde{\gamma}_1, z) = p(\tilde{\gamma}_2, z)$ implies $\tilde{\gamma}_1 = \tilde{\gamma}_2$ and $\tilde{\beta}_1^T z = \tilde{\beta}_2^T z$ implies $\tilde{\beta}_1 = \tilde{\beta}_2$. 
1.3 Observed Likelihood

Consider the case of right-censored data. Let $T_i$ denote the survival time of the $i$th subject and $C_i$ the random censoring time. Given $Z = z_i$, we assume that $T_i, C_i$ and $\eta_i$ are mutually independent. However, in practice we may not observe the $T_i$’s due to censoring, but instead observe $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. Thus, we observe the triplet $\{X_i, \Delta_i, Z_i\}$ which are assumed to be independent across $i = 1, \ldots, n$. We denote the set of observed data as $O = \{X_i, \Delta_i, Z_i, i = 1, \ldots, n\}$.

Hence the observed likelihood based on right-censored data can be written as

$$L(\tilde{\gamma}, \tilde{\beta}; S, O) = \prod_{i=1}^{n} \left\{ f(x | \tilde{\beta}, \tilde{\gamma}, z) \right\}^{\delta_i} \left\{ S(x | \tilde{\beta}, \tilde{\gamma}, z) \right\}^{1-\delta_i}, \quad (1.6)$$

where $f(x | \tilde{\beta}, \tilde{\gamma}, z)$ and $S(x | \tilde{\beta}, \tilde{\gamma}, z)$ are defined above in Equations (1.4) and (1.5), respectively.

The corresponding observed log likelihood is

$$l(\tilde{\gamma}, \tilde{\beta}; S, O) = \sum_{i=1}^{n} \delta_i \log \left\{ f(x | \tilde{\beta}, \tilde{\gamma}, z) \right\} + (1 - \delta_i) \log \left\{ S(x | \tilde{\beta}, \tilde{\gamma}, z) \right\}. \quad (1.7)$$

1.4 Estimation

Because Chapters 2 and 4 will consider semiparametric estimation via Bayesian and frequentist methods, respectively, we will first look at both methods in a parametric context. Estimation strategies follow.

1.4.1 Maximum Likelihood

Because the entire cure model in Equation (1.3) is parametric, the observed log-likelihood may be directly (numerically) maximized to obtain the maximum likelihood estimator (MLE) for the unknown parameters $\gamma$ and $\beta$. Further, the Fisher Information matrix can be used to obtain the asymptotic variance of the MLE. We implement this by using the “mle” function available in the “stats4” package of R. Also, the Wald-type regularity conditions are satisfied for the chosen parametric family and hence, the consistency of the asymptotic normality of the MLE’s of $\gamma$ and $\beta$ follow (Bickel and Doksum, 2001; Schervish, 1995).
1.4.2 EM Algorithm

The EM algorithm is a method used for finding maximum likelihood estimates, when the model depends on unobserved latent variables. For a semiparametric AFT cure model with no distributional assumptions on the error, Li and Taylor (2002) and Zhang and Peng (2007) both implement the EM algorithm to estimate the unknown parameters $\tilde{\gamma}$ and $\tilde{\beta}$. Though maximum likelihood estimation is straightforward, we also implement the EM algorithm. We verify reliable estimates in this simpler parametric case since one idea for future work is to develop a classical semiparametric estimation scheme that would utilize a similar EM methodology. We consider the latent variable, $\eta_i$ representing a subject’s susceptibility to experiencing a failure. The complete data consists of $(t_i, \delta_i, z_i, \eta_i)$, $i = 1, \ldots, n$. The complete-data likelihood is then given by

$$L_c(\tilde{\gamma}, \tilde{\beta}, S_c; O, \eta) = \prod_{i=1}^{n} \left\{ p(\tilde{\gamma}, z_i) \right\}^{\eta_i} \left\{ 1 - p(\tilde{\gamma}, z_i) \right\}^{1 - \eta_i} \times \left\{ f_c(\log t_i - \tilde{\beta}' \tilde{z}_i)/t \right\}^\delta \eta_i \left\{ S_c(\log t_i - \tilde{\beta}' \tilde{z}_i) \right\}^{(1 - \delta) \eta_i}$$ (1.8)

Taking the logarithm and only considering terms that depend on $\tilde{\gamma}$ or $\tilde{\beta}$ where the hazard function is denoted by $\lambda_c(\cdot) = f_c(\cdot)/S_c(\cdot)$, the complete log likelihood is $l_c(\tilde{\gamma}, \tilde{\beta}, S_c; O, \eta) = l_{c1}(\tilde{\gamma}; O, \eta) + l_{c2}(\tilde{\beta}; S_c; O, \eta)$, where

$$l_{c1}(\tilde{\gamma}; O, \eta) = \sum_{i=1}^{n} \eta_i \log \{ p(\tilde{\gamma}, z_i) \} + (1 - \eta_i) \log \{ 1 - p(\tilde{\gamma}, z_i) \}$$ (1.9)

$$l_{c2}(\tilde{\beta}; S_c; O, \eta) = \sum_{i=1}^{n} \eta_i \delta_i \log \{ \lambda_c(\log t_i - \tilde{\beta}' \tilde{z}_i) \} + \eta_i \log \{ S_c(\log t_i - \tilde{\beta}' \tilde{z}_i) \}$$ (1.10)

The EM algorithm is an iterative process with two repeating steps. First, the E-step consists of computing the expectation of the complete-data log-likelihood with respect to the conditional distribution of the latent variable $\eta$ given the current parameter estimates $\Theta^{(m)} = \{ \tilde{\beta}^{(m)}, \tilde{\gamma}^{(m)} \}$ and the observed data $O$. Let $u_i^{(m)} = E(\eta_i | \Theta^{(m)}, O)$ denote the conditional probability that the $i$th subject is in the susceptible group at the $m$th iteration of the algorithm. By a simple application of Bayes’ theorem,

$$u_i^{(m)} = \delta_i + (1 - \delta_i) \frac{p(\tilde{\gamma}, z_i)S_c(\log t_i - \tilde{\beta}' \tilde{z}_i)}{1 - p(\tilde{\gamma}, z_i) + p(\tilde{\gamma}, \tilde{z}_i)S_c(\log t_i - \tilde{\beta}' \tilde{z}_i)} \bigg|_{\Theta^{(m)}, O}$$ (1.11)
Essentially, the E-step replaces the $\eta_i$’s with $w_i^{(m)}$. Hence

$$E\{l_c(\tilde{\gamma}, \tilde{\beta}, S_\epsilon; O, \eta)\} = l_{c1}(\tilde{\gamma}; w^{(m)}) + l_{c2}(\tilde{\beta}, S_\epsilon; w^{(m)}),$$

where

$$l_{c1}(\tilde{\gamma}; w^{(m)}) = \sum_{i=1}^{n} w_i^{(m)} \log \{p(\tilde{\gamma}, z_i)\} + (1 - w_i^{(m)}) \log \{1 - p(\tilde{\gamma}, z_i)\} \quad (1.12)$$

$$l_{c2}(\tilde{\beta}, S_\epsilon; w^{(m)}) = \sum_{i=1}^{n} w_i^{(m)} \delta_i \log \{\lambda_\epsilon (\log t_i - \tilde{\beta} \cdot \tilde{z}_i)\} + w_i^{(m)} \log \{S_\epsilon (\log t_i - \tilde{\beta} \cdot \tilde{z}_i)\} \quad (1.13)$$

Secondly, the M-Step is to maximize Equations (1.12) and (1.13) with respect to the unknown parameters $\tilde{\gamma}$ and $\tilde{\beta}$. With no distributional assumptions on the error term, estimation is difficult. Li and Taylor (2002) use a score function approximation based on work by Ritov (1990) and an estimator based on Taylor (1995) applied to residuals. Recently, Zhang and Peng (2007) have developed estimators by using the Newton Raphson algorithm and making use of a rank-based estimator similar to Wei (1992).

Because of our parametric assumptions on the error distribution, the EM algorithm is easily implemented by directly maximizing the expected value of the two pieces of the complete-data log-likelihood in Equations (1.12) and (1.13). Louis (1982) has developed a procedure to compute the observed information matrix when the EM algorithm is used. Oakes (1999) developed a direct calculation of the information matrix via the EM algorithm. We implement Oakes’ method for calculation of standard errors associated with the parameter estimates, because of its simpler ease in computation. Standard error derivations are shown in Appendix A.1.

### 1.4.3 Bayesian Methodology

We specify prior distributions for Bayesian estimation. We consider priors for $\beta$ and $\gamma$ with large variance, and use $\tilde{\beta} \sim N(0, D_{\tilde{\beta}})$ and $\tilde{\gamma} \sim N(0, D_{\tilde{\gamma}})$ where each of the variances $D_{\tilde{\beta}}$ and $D_{\tilde{\gamma}}$ are adjusted to provide reasonable parameter spaces for each of the incidence and latency terms. We specify a parametric error distribution, in which case the $S_\epsilon(\cdot)$ has known parametric forms with unknown scale term $\sigma_\epsilon^2$. A noninformative prior is set for the scale, $\sigma_\epsilon^2 \sim IG(a_0, b_0)$ where $IG(\cdot, \cdot)$ denotes an inverse-gamma distribution with $a_0 > 0$ and $b_0 > 0$ suitably chosen to obtain a prior with large variance. Define the set of parameters to be estimated as $\theta = \{\tilde{\beta}, \tilde{\gamma}, \sigma_\epsilon^2\}$. Having specified the observed likelihood in
Equation (1.6) and designated appropriate prior distributions, the posterior distribution of the parameters $\theta$ is

$$\pi(\theta | O) \propto L(\tilde{\gamma}, \tilde{\beta}, \sigma^2; O) \pi(\sigma^2) \pi(\tilde{\beta}) \pi(\tilde{\gamma}). \quad (1.14)$$

The posterior distribution of the parameters can not be obtained in analytical closed form. Instead, MCMC sampling from the marginal posterior distributions of $\tilde{\beta}$ and $\tilde{\gamma}$ derived from Equation (1.14) is performed. The posterior median and posterior standard deviations are used for measures of center and spread. We implement this by using the “bugs” function available in the “R2WinBUGS” package of R which calls WinBUGS to generate samples from the posterior distributions using MCSC methods. Also, appealing to the Bernstein-von Mises Theorem, the consistency of the MLE in this case as well as prior positivity imply posterior consistency (Bickel and Doksum, 2001)

### 1.5 Simulation Studies

We conduct several simulation studies to analyze how well the MLE, EM algorithm-based estimators, and estimators via Bayesian posterior summaries perform under a variety of settings for finite sample sizes. We consider two covariates, $z_1$ is uniformly distributed between 0 and 1 and $z_2$ is a 0-1 binary variable from a Bernoulli distribution with mean 0.5. For the Bayesian analyses, covariates are centered by their population mean, which is 0.5 for both variables. Such centering is known to reduce the correlations between the regression coefficients and hence, leads to more efficient Markov chain Monte Carlo sampling schemes (Roberts and Sahu, 2001).

The incidence portion of the model is

$$\log[p(\tilde{\gamma}, z) / \{1 - p(\tilde{\gamma}, z)\}] = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2.$$

We set $\gamma_0 = 0.5$ or 1.0, $\gamma_1 = 1.0$, and $\gamma_2 = -1.0$. Cure fraction and censoring percentages correspond to 38% and 43% when $\gamma_0 = 0.5$ and 28% and 34% when $\gamma_0 = 1.0$. The latency portion of the model is

$$\log T^* = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \epsilon$$

where $\beta_0 = -1$, $\beta_1 = -1.0$ and $\beta_2 = 1.0$. Data are generated from the cure model in Equations (1.1)-(1.2) according to three error distributions: (i) Normal(0,1) with density function $g(\epsilon; a, b) = (2\pi\sigma^2)^{-1/2} \exp\{(-\epsilon - \mu)^2/(2\sigma^2)\}$, (ii) EV(0,1) where EV represents the extreme value distribution with density function $g(\epsilon; a, b) = b^{-1} \exp\{(-\epsilon/a) \exp[-\exp((\epsilon - a)/b)]\}$, and (iii) Logistic(0,1) with density function $g(\epsilon; a, b) = [1 + \exp\{-(\epsilon-a)/b\}]^{-1}$. The Bayesian estimation only considers error densities from (i) and (iii) whereas the frequentist methods consider all three cases.
The censoring times are generated from a uniform(0, 10) distribution. The observed times are hence the minimum of the failure and censoring times. We use sample sizes of \( n = 100 \) for the frequentist analyses and sample sizes of either \( n = 100 \) or 200 subjects for the Bayesian analyses. Estimation of the incidence parameters which control the tail of the survival curve can be difficult with occasional inflated estimates. For frequentist analyses, when \( \gamma_0 \) estimates were greater than 4, the data sets were dropped. This only occurred with the logistic distribution in 1-2 instances out of 500 replications.

In terms of prior specification for the Bayesian estimation, originally we set \( \beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1, \gamma_2 \sim N(0, 10) \). Because the \( \gamma_0 \) estimates tends to suffer from large biases, we then use a slightly more informative prior for \( \gamma_0 \): \( \gamma_0 \sim N(0, 9) \). Also, \( a_0 = 0.1 \) and \( b_0 = 10 \) so \( \sigma^2 \sim Ga(0.1, 10) \). For each Bayesian simulation, the posterior generation in WinBUGS is based on 2000 burn-ins and an additional 5000 runs from three parallel chains, for a total of 15,000 Markov chain Monte Carlo samples per simulated data.

For all simulation scenarios, 500 Monte Carlo samples are computed. For frequentist analyses, descriptive statistics include the bias associated with the mean of the parameter estimates, the sample standard error of the estimators (MCSE), the average of the model-based estimated standard errors (ESE), the mean-squared error (MSE), and the empirical coverage probabilities for 0.95 confidence intervals. Standard error estimates associated with the EM algorithm-based estimators are computed by direct calculation based on Oakes’ formula. For Bayesian analyses, the bias associated with the posterior median, the Monte Carlo standard error (MCSE) of the posterior median, the MC average of the posterior standard deviation (ESE), and the nominal coverage probabilities of covering the true value based on the equal-tailed 95% posterior interval.

Tables A.2-A.4 in Appendix A.3 provides results for Monte Carlo simulation studies based on frequentist estimation strategies. For the ML and EM-based estimators, the empirical means of the estimates all seem to be fairly close to their true values. The biases for the two estimation techniques are nearly identical. The averages of the estimated standard errors associated with the MLE are closer to the Monte Carlo standard errors than those associated with the EM algorithm. Because of this discrepancy, the coverage probabilities for ML estimation are closer to the nominal level.

In all the ML and EM simulations, the standard errors associated with \( \gamma_1 \) seem to be a little large, suggesting slight inflation due to estimation difficulties in the tail end of the
survival curve. This is also visible in the boxplots, but may be corrected by adjusting the data inclusion criterion to include a cutoff estimation value for $\gamma_2$. The coverage probabilities associated with the beta estimates for the normal distribution, and that for $\beta_1$ in the logistic case are a little smaller than expected. Otherwise, the estimators seem to be performing well.

Table A.5 in Appendix A.3 contains additional simulations based on a logistic(0,7) error term for the ML and EM-based estimation. The variance associated with a logistic(0,b) density is $(b \pi^2)/3$, where $b$ is the scale factor. The logistic(0,1) density has a variance of roughly 3.3 whereas the logistic(0,7) density has a variance of roughly 1.6. The low coverage probabilities in the logistic(0,1) simulation seem to be related to the long tails of the distribution. By shortening the tails in the logistic(0,7) case, the coverage probabilities have improved and are more in line with the normal and extreme value simulations.

Appendix A.4 provides the Bayesian estimation results. Table A.6 provides results to the first set of Monte Carlo simulation studies based on Bayesian estimation methods. The error terms are generated from standard normal and logistic densities and all priors are noninformative. With the exception of $\gamma_0$ in the logistic error case, the posterior medians of all parameters seem to be performing comparably to the maximum likelihood and EM algorithm estimation methods. The coverage probabilities are close to their nominal values. With the exception of the biased $\gamma_0$ estimate in the logistic error case, overall the simulations seems to be performing well.

Tables A.7-A.8 provides results to the eight Monte Carlo simulations that use a slightly more informative prior for $\gamma_0$ and consider an increased sample size and increased cure fraction. When $n = 100$, introducing a more informative prior results in a small reduction in the bias of $\hat{\gamma}_0$ for both the standard normal and logistic errors. And when the sample size is increased to $n = 200$, the bias is finally reduced to a manageable level even in the logistic case. The bias of $\hat{\gamma}_0$ seems to be slightly less in each case when $\gamma_0 = 0.5$ rather than 1.0.

1.6 Conclusion

We have provided both frequentist and Bayesian parametric estimation procedures for the CRAFT model and verified that each provide good estimates of both the latency and
incidence parameters. However, the underlying error distribution is rarely known, so it is of interest to provide an alternative method which relaxes this assumption. Chapter 2 will provide an alternative semiparametric CRAFT model whereby a mixture distribution is used to model the error. Bayesian estimation methods are employed to obtain estimates. In a frequentist context, both Li and Taylor (2002) and Zhang and Peng (2007) have proposed semiparametric methods by working with the complete log-likelihood and implementing the EM algorithm to jointly maximize $\beta$ and $\gamma$. Due to the difficulty in the simultaneous maximization of the $\beta$ term along with both the error density $f(\cdot)$ and survival function $S(\cdot)$, neither provides true maximum likelihood estimates. As an alternative estimation method, the semiparametric CRAFT model which will be introduced in Chapter 2 could be used. Then a semiparametric EM-type approach may be implemented to provide maximum likelihood estimates of both the incidence and latency parameters.
Chapter 2

The Semiparametric CRAFT model Under Right Censoring

2.1 Introduction

Semiparametric estimation methods are becoming increasingly popular for both the accelerated failure time (AFT) model as well as cure rate AFT (CRAFT) model. It is advantageous to avoid making possibly incorrect assumptions, especially when it is rare in practice to actually know the underlying error distribution associated with the model. So there is interest in revamping the parametric CRAFT model in Chapter 1 by relaxing the known error distribution assumption.

With no error distribution assumption, simultaneous estimation of both incidence and latency regression parameters as well as the underlying survival distribution becomes increasingly more difficult for CRAFT models. In a frequentist context, Li and Taylor (2002) and Zhang and Peng (2007) both employ the EM algorithm to reduce the dimension of estimation by separating out the incidence and latency portions of the CRAFT model. In a Bayesian context, Chen et al. (1999), Ibrahim et al. (2001), Yakovlev and Tsodikov (1996), and Kim et al. (2007) have proposed semiparametric estimation procedures, but using the promotion time cure model rather than the two-component mixture cure model framework used in the CRAFT model. None of these papers provide large sample theoretical justification. The AFT model for general right censoring (without a cure fraction) has been studied by several including Christensen and Johnson (1988), Johnson and Christensen
(1989), Kuo and Mallick (1997), Walker and Mallick (1999), Campolieti (2001), and Hansen and Johnson (2004). Ghosh and Ghosal (2006) establish posterior consistency based on a class of mixture of Weibull densities. Wu (2009) has extended their result by relaxing the assumption that the true error density itself is a mixture of weibull mixtures. However, to the best of our knowledge, in a Bayesian context posterior consistency has not been established for an AFT model with a cure fraction.

We propose a semiparametric CRAFT model with smoothed error distribution for right-censored data. The error distribution is modeled as an infinite mixture of normal densities, allowing for nonparametric estimation in the latency. Normal mixtures have been explored in the statistical literature by many authors, including Teicher (1961), Redner and Walker (1984), Ahmad (1988), Roeder (1992), Magder and Zeger (1996), and Komárek et al. (2005). With suitable prior specifications, Markov chain Monte Carlo methodology is used to obtain estimates based on posterior distributions. Using similar ideas as in Ghosh and Ghosal (2006) for the standard accelerated failure time setting, certain compactness assumptions on the parameter spaces are used to rigorously establish consistency of the posterior distribution. Simulations studies evaluate the performance of the model and the estimation methods are illustrated with a breast cancer data set.

2.2 The CRAFT Model with Unspecified Error Distribution

Recall $T$ is the time to occurrence of event which can be represented as $T = T^*\eta + \infty(1 - \eta)$ where $T^*$ denotes the latent survival time when subjects are susceptible. The incidence is still modeled with the logistic link function in Equation (1.1), as

$$P(\eta = 1 \mid z) = p(\tilde{\gamma}, z) = F_0(\tilde{\gamma}^T \tilde{z}).$$

The latency portion of the model does not include an intercept term, so the AFT component to the CRAFT model from Equation (1.2) is now

$$\log T^* = \beta^T z + \epsilon,$$

where $\beta$ is a $p$-dimensional vector of unknown incidence parameters, $\epsilon$ represents measurement error, and $\text{var}(\epsilon) = \sigma^2_\epsilon$. Again, we assume that the same vector of covariates $z$ is present in the incidence and latency components. Also, we assume that the error $\epsilon$ follows
an infinite mixture of normals with an unknown mixing distribution $H(\cdot)$, where $H(\cdot)$ is an unknown cumulative distribution function satisfying

$$
\int (\mu - \mu_\epsilon)^2 dH(\mu) = \left( \frac{k_0}{k_0 + 1} \right) \sigma_\epsilon^2,
$$

(2.2)

$\mu_\epsilon = \int \mu dH(\mu)$, and $k_0 > 0$ is chosen arbitrarily. For identifiability we assume that $k_0$ is fixed and can be suitably chosen. In other words, the probability density function, $g(\epsilon)$ of $\epsilon$ is given by

$$
g(\epsilon) = \int \frac{(k_0 + 1)^{1/2}}{\sigma_\epsilon} \phi \left( \frac{\epsilon - \mu}{\sigma_\epsilon/(k_0 + 1)^{1/2}} \right) dH(\mu),
$$

(2.3)

where $\phi(\cdot)$ is the standard normal density. Notice that the density in Equation (2.3) satisfies the condition $\text{var}(\epsilon) = \sigma_\epsilon^2$ for any cumulative density function $H(\cdot)$ and $k_0 > 0$. We propose suitable priors for $\beta, \tilde{\gamma}, H(\cdot)$, and $\sigma_\epsilon^2$ and make additional assumptions to establish posterior consistency. Having defined both the incidence and latency, the general cure rate accelerated failure time model is

$$
P(T > t \mid Z = z) = S(t \mid \beta, \tilde{\gamma}, H(\cdot), \sigma_\epsilon^2, z) = 1 - p(\tilde{\gamma}, z) + p(\tilde{\gamma}, z) P(T^* > t \mid Z = z),
$$

(2.4)

where

$$
P(T^* > t \mid Z = z) = S_{T^*}(t \mid \beta, \tilde{\gamma}, H(\cdot), \sigma_\epsilon^2, z) = \int \left\{ 1 - \Phi \left( \frac{\log t - \mu - \beta^T z}{\sigma_\epsilon/(k_0 + 1)^{1/2}} \right) \right\} dH(\mu),
$$

(2.5)

and $\Phi(\cdot)$ is the standard normal cumulative distribution function. The corresponding density function for survival times $T = t$ is

$$
f(t \mid \beta, \tilde{\gamma}, H(\cdot), \sigma_\epsilon^2, z) = p(\tilde{\gamma}, z) \int \phi \left( \frac{\log t - \mu - \beta^T z}{\sigma_\epsilon/(k_0 + 1)^{1/2}} \right) \left( \frac{k_0 + 1)^{1/2}}{\sigma_\epsilon t} dH(\mu) \quad \text{for} \ t < \infty.
$$

(2.6)

### 2.3 Implementation

In practice, the mixing distribution can often be approximated by a finite, possibly sample size-dependent, discrete distribution (Li and Barron, 2000; Komárek et al., 2005). We approximate the error density in Equation (2.3) by

$$
g_{L_n}(\epsilon) = \sum_{l=1}^{L_n} w_l \left( \frac{k_0 + 1)^{1/2}}{\sigma_\epsilon} \phi \left( \frac{\epsilon - \mu_\epsilon}{\sigma_\epsilon/(k_0 + 1)^{1/2}} \right),
$$

(2.7)
where $L_n$ is the number of normal mixtures, $\mu_1 < \cdots < \mu_{L_n}$ is a suitably chosen completely known ordered sequence of knots in $\mathbb{R}$, and $w = (w_1, \ldots, w_{L_n})^T$ are unknown non-negative mixture coefficients satisfying the restrictions (i) $\sum_{l=1}^{L_n} w_l = 1$ and (ii) $\sum_{l=1}^{L_n} \mu_l^2 w_l - (\sum_{l=1}^{L_n} \mu_l w_l)^2 = \{k_0/(k_0 + 1)\} \sigma^2$, where $k_0$ is a suitably chosen positive number. The first restriction guarantees that $g_{L_n}(\epsilon)$ is a density function while the second restriction ensures that the variance constraint from Equation (2.2) holds in the finite mixture. We define $L$ with a subscript $n$ to reflect the dependency between the number of normal densities and the sample size, which will be discussed in more detail in section 2.4.

Komárek et al. (2005), building on work by O’Sullivan (1986) and Eilers and Marx (1996), selected a relatively large but fixed number of equidistant knots in a finite normal mixture. The flexibility of the curve was restricted by including a penalty term based on squared finite higher-order differences between adjacent mixture coefficients within the likelihood function. A fine grid of knots ensured accurate estimation of the error density while the penalty term avoided over-fitting. Alternatively, we develop suitable prior distributions for $w$ which achieves a similar penalization within our Bayesian framework.

If we choose to use a suitable Dirichlet distribution to model the weight vector $w$ then Equation (2.7) provides an approximation to a mixture of Dirichlet process priors (Ishwaran and Zarepour, 2002). However, in this paper we develop a more flexible prior for $w$ which is required to satisfy the additional variance constraint for model identifiability. We represent the $w_i$’s as a multivariate logit model; as

$$w_l = \frac{e^{\alpha_l}}{\sum_{l=1}^{L_n} e^{\alpha_l}} \quad (l = 1, \ldots, L_n - 1),$$

$$w_{L_n} = 1 - \sum_{l=1}^{L_n-1} w_l = \frac{\sum_{l=1}^{L_n-1} e^{\alpha_l}}{1 + \sum_{l=1}^{L_n-1} e^{\alpha_l}},$$

where $\alpha_{L_n} = 0$ and $\alpha_l \in \mathbb{R}$ for $l = 1, \ldots, L_n - 1$. Thus there is no restriction on the $\alpha_l$’s for $l = 1, \ldots, L_n - 1$ and this parameterization allows restriction (i) above to inherently hold.

We derive a prior for $\alpha = (\alpha_1, \ldots, \alpha_{L_n-1})^T$ which in turn induces a prior on $w$ in a similar spirit as the penalty term used by Komárek et al. (2005). We assume that $\alpha \sim N_{L_n-1}(0, \lambda^{-1}(D_m^T D_m)^{-})$ where $D_m$ is a $(L_n - m - 1) \times (L_n - 1)$ matrix to represent $m$th order differences (see Appendix B.1 for more details) and $(D_m^T D_m)^{-}$ is the Moore-Penrose generalized inverse. Notice that $D_m$ is a matrix of rank $(L_n - m - 1)$ and hence, $\alpha$ has a singular normal distribution (Rao, 1973). But if we define $\tau = D_m \alpha$ then $\tau_i \overset{iid}{\sim} N(0, \lambda^{-1})$
for \( l = 1, \ldots, L_n - m - 1 \).

Finally, we consider noninformative priors for \( \beta \) and \( \tilde{\gamma} \), and use \( \beta \sim N(0, D_\beta) \) and \( \tilde{\gamma} \sim N(0, D_{\tilde{\gamma}}) \) where each of the variances \( D_\beta \) and \( D_{\tilde{\gamma}} \) are adjusted to provide reasonable parameter spaces for each of the incidence and latency terms. Also, a noninformative prior is set for the error variance, \( \sigma_\epsilon^2 \sim IG(a_0, b_0) \) where \( a_0 > 0 \) and \( b_0 > 0 \) are suitably chosen to obtain a vague prior. For the rest of the paper we define the set of parameters to be estimated as \( \theta = \{ \beta, \tilde{\gamma}, \sigma_\epsilon^2, \tau \} \in \{ \mathbb{R}^p \times \mathbb{R}^{p+1} \times (0, \infty) \times \mathbb{R}^{L-m-1} \} \).

### 2.4 Semiparametric Estimation Using the Posterior Distribution

Let \( T_i \) denote the survival time of the \( i \)th subject and \( C_i \) the random censoring time. Given \( Z = z_i \), we assume that \( T_i \) is independent of \( C_i \). However, in practice we may not observe the \( T_i \)'s due to censoring, but instead observe \( X_i = \min(T_i, C_i) \) and \( \Delta_i = I(T_i \leq C_i) \). Thus, we observe the triplet \( \{X_i, \Delta_i, Z_i\} \) which are assumed to be independent across \( i = 1, \ldots, n \). We denote the set of observed data as \( O = \{X_i, \Delta_i, Z_i, i = 1, \ldots, n\} \). Then using the finite discrete distribution for the mixing distribution, the corresponding density and survival functions for the \( i \)th observation of the observed data given the parameter \( \theta \) are

\[
f_{L_n}(x_i \mid \theta, z_i) = p(\tilde{\gamma}, z_i) \sum_{l=1}^{L_n} w_l (k_0 + 1)^{1/2} \frac{\phi\left( \frac{\log x_i - \mu_l - \beta^T z_i}{\sigma_\epsilon/(k_0 + 1)^{1/2}} \right)}{\sigma_\epsilon x_i} \tag{2.8}
\]

and

\[
S_{L_n}(x_i \mid \theta, z_i) = 1 - p(\tilde{\gamma}, z_i) \sum_{l=1}^{L_n} w_l \Phi\left( \frac{\log x_i - \mu_l - \beta^T z_i}{\sigma_\epsilon/(k_0 + 1)^{1/2}} \right). \tag{2.9}
\]

Hence the observed likelihood based on right-censored data can be written as

\[
L(\theta; O) = \prod_{i=1}^{n} \left\{ f_{L_n}(x_i \mid \theta, z_i) \right\}^{\delta_i} \left\{ S_{L_n}(x_i \mid \theta, z_i) \right\}^{1-\delta_i}, \tag{2.10}
\]

where \( f_{L_n}(\cdot \mid \theta, z) \) and \( S_{L_n}(\cdot \mid \theta, z) \) are defined above in Equations (2.8) and (2.9), respectively. Having specified the observed likelihood in Equation (2.10) and prior distributions in the previous section, the posterior distribution of the parameters \( \theta \) is

\[
\pi(\theta \mid O) \propto L(\theta; O) \pi(\beta)\pi(\tilde{\gamma})\pi(\sigma_\epsilon^2)\pi(\tau).
\]
For this proposed cure rate accelerated failure time model, the posterior distribution of the parameters cannot be obtained in closed form. Using the finite discrete approximation to the infinite normal mixture allows WinBUGS to perform Markov chain Monte Carlo sampling from the posterior distribution. As shown explicitly in Ghosh and Ghosal (2006), latent variables \( L^* = (L_1^*, \ldots, L_n^*) \) which indicate the group membership to the \( \mu_l \) node along with probability vector \( w = (w_1, \ldots, w_{L_n})^T \) are introduced.

The cure rate accelerated failure time model can be re-written in hierarchical format as the following:

\[
\begin{align*}
\log T_i \mid L_i^*, \eta_i & \sim \begin{cases} 
\text{Normal} \{ \beta^T z_i + \mu_{L_i^*}, \sigma^2_i/(k_0 + 1) \}, & \delta_i = 1, \eta_i = 1 \\
\text{Normal} \{ \beta^T z_i + \mu_{L_i^*}, \sigma^2_i/(k_0 + 1) \} I(\log x_i, \infty), & \delta_i = 0, \eta_i = 1
\end{cases} \\
\eta_i \mid \tilde{z}_i & \sim \text{Bernoulli} \{ p(\tilde{\gamma}, z_i) \} \\
L_i^* \mid w & \sim \text{Multinomial} \{ (1, \ldots, L_n), w \} \\
w & = \text{Multivariate Logit}(\alpha), \quad (\tau = D_m \alpha) \\
\tau & \sim \text{Normal} \{ 0, \lambda^{-1} I_{L_n} \} \\
\beta & \sim \text{Normal} \{ 0, D_\beta \} \\
\tilde{\gamma} & \sim \text{Normal} \{ 0, D_{\tilde{\gamma}} \} \\
\sigma^2_i & \sim \text{Inverse Gamma}(a_0, b_0)
\end{align*}
\]

Note that \( \text{pr}(L_i^* = l) = w_l \) for \( l = 1, \ldots, L_n \). This hierarchical format is easier to implement in WinBUGS. We select knots ranging from \( \mu_1 = -M \) to \( \mu_{L_n} = M \) where \( \mu_l = -M + 2M(l - 1)/(L_n - 1), l = 1, \ldots, L_n \) for some suitably chosen large \( M > 0 \). This should be a wide enough range to account for densities that may have large tails, such as the extreme value or logistic distributions. Ishwaran and Zarepour (2002) suggest using \( L_n = n^{1/2} \) for large \( n \) and \( L_n = n \) for small \( n \) while Ghosh and Ghosal (2006) say that more work is necessary to determine an optimal number. We will select an order of \( L_n \) in a similar fashion, that should allow each normal density to overlap with a few of its neighborhoods and may increase with \( n \). Alternatively, a reversible jump MCMC procedure which has recently become available in WinBUGS could have been used to select \( L_n \) (Lunn et al., 2008). Selecting \( m = 2, 3, \) or \( 4 \) in the difference operator matrix seems to provide good smoothing of the density in simulations. Also, we set \( \lambda = 1 \) but in the future it may be determined by a cross-validation procedure suggested by Komárek et al. (2005) or estimated
using a prior distribution.

With this hierarchical formulation of our model specifications, Markov chain Monte Carlo sampling may be easily performed to generate samples from the marginal posterior distributions of each of the parameters. The posterior mean and standard deviation are used for the measure of center and spread for the mixture coefficients, while the posterior median and standard deviation are used for all other estimated parameters. We implement this by using the “bugs” function available in the “R2WinBUGS” package of R which calls WinBUGS to generate samples from the posterior distributions using Markov chain Monte Carlo methods. Convergence diagnostics were performed using the “CODA” package available in R. See Appendix B.3 for the associated WinBUGS code.

2.5 Consistency of Posterior Distribution

Consistency is a desirable large sample property of the posterior distribution. As Ghosal (2000) discusses, it guarantees that the posterior distribution will concentrate in arbitrarily small neighborhoods of the true value of the parameter, and hence with a sufficiently large amount of data, the truth may be discovered accurately. Diaconis and Freedman (1986) and Ghosh and Ramamoorthi (2003) provide nice discussions and examples of posterior consistency. A formal definition of posterior consistency is as follows.

**Definition (Posterior Consistency).** Suppose $X_1, X_2, \ldots$ are independent and identically distributed according to an unknown density $f_0$. We take the parameter space as $\mathcal{F}$ - a set of probability densities on the space of the observations and consider a prior distribution $\Pi$ on $\mathcal{F}$. Then the posterior distribution $\Pi(\cdot | X_1, \ldots, X_n)$ of $f \in \mathcal{F}$ given a sample $X_1, \ldots, X_n$ is obtained as,

$$
\Pi(A | X_1, \ldots, X_n) = \frac{\int_A \prod_{i=1}^{n} f(X_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^{n} f(X_i) d\Pi(f)}.
$$

We say that the posterior achieves weak posterior consistency at $f_0$ if for any weak neighborhood $U$ of $f_0$, $\Pi(U | X_1, \ldots, X_n) \to 1$ almost surely as $n \to \infty$.

Sufficient conditions for posterior consistency involving appropriate tests and the prior positivity of a neighborhood defined by the Kullback-Leibler divergence are presented in a theorem by Schwartz (1965). Barron et al. (1999), Ghosal et al. (1999a), Ghosh and Ramamoorthi (2003), and Amewou-Atisso et al. (2003) explore useful extensions of Schwartz’s
Theorem for density estimation in mixture models with and without covariates. In-depth justifications and proofs of consistency for the semiparametric accelerated failure time model with infinite Weibull density mixture using Dirichlet priors for the mixing distribution have been presented by Ghosh and Ghosal (2006). We follow Ghosh and Ghosal’s methodology to show consistency for our CRAFT model with infinite normal mixture and hence provide a large sample justification of our Bayesian analysis.

The following assumptions are sufficient to establish posterior consistency. We assume that the domains of $Z, \beta, \tilde{\gamma}$, and $\sigma^2_t$ and the support of $H$ are compact. We assume that the zero vector is a possible value of the covariate $Z$, which if not, the covariates may be shifted to satisfy this condition. We assume that the true density $f^*$ of $T$ given $Z = z$ is a mixture of normal densities

$$f^*(t \mid z) = p(\tilde{\gamma}_*, z) \int_0^\infty \frac{(k_0 + 1)^{1/2}}{\sigma_* t} \phi \left( \frac{\log t - \beta_*^t z - \mu}{\sigma_* / (k_0 + 1)^{1/2}} \right) dH_*(\mu)$$

where $\beta_*, \tilde{\gamma}_*, \sigma^2_*$, and $H_*$ are the true values of the parameters $\beta, \tilde{\gamma}, \sigma^2_*$, and $H$, respectively. Notice that the above assumption can be relaxed using the recent results obtained by Wu and Ghosal (2008). However, since a mixture of normal densities can be used to approximate any bounded continuous density in total variation norm, the assumption about $f^*(t \mid z)$ is not too restrictive. We have fixed, independently distributed variables, with absolutely continuous distribution supporting the vector 0 in $\mathbb{R}^p$. Denote the density of $Z$ at $z$ by $q(z)$. Let $h_{\beta, \tilde{\gamma}, \sigma^2_*, H}(x, \delta, z)$ be the joint density of $(X, \Delta, Z)$, so

$$h_{\beta, \tilde{\gamma}, \sigma^2_*, H}(x, \delta, z) = \begin{cases} 
  f\{x \mid \beta, \tilde{\gamma}, H(\cdot), \sigma^2_*, z\} q(z), & \delta = 1, \\
  S\{x \mid \beta, \tilde{\gamma}, H(\cdot), \sigma^2_*, z\} q(z), & \delta = 0,
\end{cases}$$

(2.11)

where $f\{x \mid \beta, \tilde{\gamma}, H(\cdot), \sigma^2_*, z\}$ and $S\{x \mid \beta, \tilde{\gamma}, H(\cdot), \sigma^2_*, z\}$ are defined in Equations (2.4)–(2.6). The class of distributions that are supported in a given compact domain is also compact with respect to the weak topology on the space of probability measures. So the parameter space of $(\beta, \tilde{\gamma}, \sigma^2_*, H)$ with respect to the product of Euclidean and weak topology is also compact. Hence, the following main theorem applies which verifies that the posterior distribution is consistent.

**Theorem 1.** Suppose that the prior densities $\pi(\beta), \pi(\tilde{\gamma}),$ and $\pi(\sigma^2_*)$ for $\beta, \tilde{\gamma},$ and $\sigma^2_*$ have compact supports containing $\beta_*, \tilde{\gamma}_*$, and $\sigma^2_*$, the base measure of $H$ has compact support
that contains the support of $H_*$, and $H_*$ satisfies the constraint in Equation (2.2). Then the posterior distribution $\Pi((\beta, \tilde{\gamma}, \sigma^2_\epsilon, H) \in \cdot \mid (X_1, \Delta_1), \ldots, (X_n, \Delta_n))$ of $(\beta, \tilde{\gamma}, \sigma^2_\epsilon, H)$ given $(X_1, \Delta_1), \ldots, (X_n, \Delta_n)$ is consistent with respect to the Euclidean distances on $\beta, \tilde{\gamma}$, and $\sigma^2_\epsilon$ and the weak topology on $H$, that is, given any $\epsilon > 0$ and a weak neighborhood $N$ of $H_*$,

$$
\Pi((\beta, \tilde{\gamma}, \sigma^2_\epsilon, H) : \\
|\beta - \beta_*| < \epsilon, |\tilde{\gamma} - \tilde{\gamma}_*| < \epsilon, |\sigma^2_\epsilon - \sigma^2_\epsilon_*| < \epsilon, H \in N \mid (X_1, \Delta_1), \ldots, (X_n, \Delta_n)) \to 1
$$

almost surely in $P^\infty_{(\beta, \tilde{\gamma}, \sigma^2_\epsilon, H_*)}$ probability.

In showing consistency we have provided a large sample justification of our Bayesian method. The proof of Theorem 1 is given in Appendix B.2.

### 2.6 Simulation Studies

We conduct several simulations to analyze how well the estimators obtained from the posterior distributions are performing under a variety of settings. We consider two covariates, $z_1$ is uniformly distributed between 0 and 1 and $z_2$ is a 0-1 binary variable from a Bernoulli distribution with mean 0.5. Covariates are centered by their population mean, which is 0.5 for both variables. Such centering is known to reduce the correlations between the regression coefficients and hence, leads to more efficient Markov chain Monte Carlo sampling schemes (Roberts and Sahu, 2001).

The incidence portion of the model is log $\{p(\tilde{\gamma}, z)/[1 - p(\tilde{\gamma}, z)]\} = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2$. We set $\gamma_0 = 0.5$ or 1.0, $\gamma_1 = 1.0$, and $\gamma_2 = -1.0$. Cure fraction and censoring percentages correspond to 39% and 42% when $\gamma_0 = 0.5$ and 28% and 32% when $\gamma_0 = 1.0$. The latency portion of the model is log $T^* = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \epsilon$ where $\beta_0 = -1$, $\beta_1 = -1.0$ and $\beta_2 = 1.0$. We consider symmetric, skewed, and bimodal distributions for $\epsilon$, each of which has mean 0 and variance 0.25. These include (i) Logistic(0, 0.28) with density function $g(\epsilon; a, b) = \exp\{-(\epsilon - a)/b\}/(b[1 + \exp\{-(\epsilon - a)/b\}]^2)$, (ii) EV(0.23, 0.39) where EV represents the extreme value distribution with density function $g(\epsilon; a, b) = b^{-1} \exp\{((\epsilon - a)/b) \exp[-\exp\{((\epsilon - a)/b)\}]\}$, and (iii) mixture of two normal densities: $N(-0.45, 0.04)$ and $N(0.45, 0.055)$, with density function $g(\epsilon; \mu_1, \mu_2, \sigma^2_1, \sigma^2_2) = (2\pi\sigma^2_1)^{-1/2} \exp\{(\epsilon - \mu_1)^2/(2\sigma^2_1)\} + (2\pi\sigma^2_2)^{-1/2} \exp\{(\epsilon - \mu_2)^2/(2\sigma^2_2)\}$. 
The censoring times are generated from a uniform$(0,10)$ distribution. The observed times are hence the minimum of the failure and censoring times. We use sample sizes of $n = 100$ or $200$ subjects. In terms of prior specification, we set $\beta_1, \beta_2, \gamma_1, \gamma_2 \sim N(0,3)$ and the incidence intercept $\gamma_0 \sim N(0,1.5)$. Also, $a_0 = 0.1$ and $b_0 = 10(k_0 + 1)$ so $\sigma^2_\epsilon \sim Ga(0.1, 10(k_0 + 1))$. The mixture of normal densities used to estimate the error term range from $\mu_l = -4, \ldots, 4$ with $L_n = 17$ nodes. We set $m = 2$ in the difference operator matrix to control the smoothing of the normal mixtures. The prior on $\tau_l$ used in the error densities is $\tau_l \sim N(0, \lambda^{-1})$ for $l = 1, \ldots, 14$ where $\lambda = 1$. For each simulation, the posterior generation in WinBUGS is based on 4000 burn-ins and an additional 2000 runs from three parallel chains, for a total of 6000 Markov chain Monte Carlo samples per simulated data.

Original simulations resulted in considerable biases associated with the $\gamma_0$ estimates, even with a more precise prior. To increase precision, we force the survival times $S_{Ln}(\cdot \mid \theta, z) = 1 - p(\tilde{\gamma}, z)$ in Equation (2.9) for censored observations with event times greater than the maximum of the uncensored failure times.

For all these simulation scenarios, 1000 Monte Carlo samples were generated to study the sampling variability of the posterior estimates of $\beta_1, \beta_2, \gamma_0, \gamma_1,$ and $\gamma_2$. In order to measure the empirical performance of these posterior estimates, we computed the bias of the posterior median, the Monte Carlo standard error of the posterior median, the Monte Carlo average of the posterior standard deviation, and the nominal coverage probabilities of covering the true value based on the 95% posterior interval.

Figure 2.1: Estimated (based on semiparametric CRAFT model) and true survival probabilities for data generated from logistic, extreme value, and bimodal distributions when $n = 200$ and $\gamma_0 = 1$ for Monte Carlo simulations of size 1000.
Tables B.1-B.3 in Appendix B.4 provides results to the Monte Carlo simulations. The posterior medians perform well in all simulations except for the $\gamma_1$ estimates, corresponding to the continuous covariate. For all error densities, these estimates tend to be biased when $n = 100$. In the higher cure fraction case, the bias disappears for $n = 200$. In the lower cure fraction cases, the bias begins to reduce at a sample size of $n = 200$ and completely disappears for larger sample size simulations not shown. Overall, the posterior medians perform well and any biases are due to small sample sizes. Coverage probabilities are close to their nominal values in all settings. The Monte Carlo standard errors of the posterior medians and Monte Carlo averages of the posterior standard deviations are fairly close and any differences reduce to a minimal amount when $n = 200$. Again, the posterior estimates seem to be performing well, with the only complication being some bias in continuous covariate effects in small sample sizes. Also, the estimated mixture densities capture the true underlying error distributions very well for all simulation scenarios. As illustration, Figure 2.1 shows the estimated and true survival distributions when $n = 200$ and $\gamma_0 = 1$ for each of the three error distributions.

2.7 Breast Cancer Data Set

Farewell (1986) analyzed a breast cancer data set to demonstrate the effectiveness of a Weibull-cure mixture model. Kuk and Chen (1992), Peng and Dear (2000), and Lu and Ying (2004) have each re-analyzed the same data set for different proportional hazards cure models. Lu and Ying (2004) also analyzed the breast cancer data set for a proportional odds cure model.

The data set consists of $n = 139$ patients, where time to relapse or death is used as the failure endpoint. Subjects are randomly assigned to one of three treatments, resulting in two treatment indicator variables. Two more binary covariates are considered: a clinical stage indicator and an indicator for the number of lymph nodes. In the original data set as presented by Farewell (1986) two more covariates were included, pathological stage and histological stage, but both have since been lost. The censoring percentage is 68%, with 95 patients censored and 44 uncensored. Figure 2.2 shows the Kaplan–Meier survival curves for each of the three treatment groups. The tails of the survival curves level-off significantly above zero for each treatment group. The failure times are shown in days on the logarithm
There is some empirical evidence that the proportional hazards assumption is not valid. This is seen by plotting the logarithm of the cumulative hazard function for the uncensored patients in each treatment group, based on the Kaplan–Meier curve, as performed by Zhang and Peng (2007) for a different data set. The uncensored subjects are assumed to be reasonably close to those that are uncured. In this instance, the logarithm of the cumulative incidence function of the uncensored subjects nearly approximates that of the uncured subjects. This is also shown in Fig. 2.2, where treatment A does not seem parallel to the other treatments. This provides empirical evidence that the Cox proportional hazards model is possibly incorrect and provides reason for fitting an alternative model.

We fit the breast cancer data set under several settings of our proposed semiparametric Bayesian model. The logarithm of the failure times are centered by either their mean, or the mean of those observations that experience a failure. This allows the normal mixture nodes to range from -4 to 4 and easily capture the spread of the data. The data is fit under all combinations of $L_n = 17$ or 25 and $m = 2, 3, \text{ or } 4$. The priors on the $w_l$ penalize high-dimensional models and help avoid overly complex models. Hence, provided that
model complexity measures remain stable, there is no need for model selection to include a second penalization for complex models as measured by DIC. Instead model selection is based on goodness of fit, measured by deviance. In terms of prior specification, we set $\beta_1, \beta_2, \gamma_1, \gamma_2 \sim N(0,3)$ and the incidence intercept $\gamma_0 \sim N(0,1.5)$. Also, $a_0 = 0.1$ and $b_0 = 10(k_0 + 1)$ so $\sigma^2_\epsilon \sim Ga(0.1, 10(k_0 + 1))$. Posterior estimation is based on 4000 burn-ins and an additional 5000 runs from three parallel chains, for a total of 15,000 Markov chain Monte Carlo samples per estimate.

Based on small levels of deviance while maintaining stable DIC values, we select the model where $L_n = 17$ and $m = 4$. The logarithm of the failure times are centered by the mean of those observations that are uncensored. The CODA package available in R is used to perform Markov chain Monte Carlo diagnostics for the chosen model. The Gelman–Rubin 97.5% shrink factors for all statistics are $\leq 1.03$, indicating good mixing and good convergence to the appropriate distributions. Appendix B.5 provides additional diagnostic information. Figure B.5 shows the trace and posterior density plots for the incidence and latency parameters while Figure B.5 provides a plot of the posterior means of the mixing coefficients. The former indicates that the chains have thoroughly mixed while the later indicates that the number and location of nodes used in the normal mixture seem to be capturing the spread of the data well.

Table 2.1: Breast cancer data: Bayesian posterior summaries for both the latency and incidence statistics associated with treatment A or B, clinical stage indication, and number of lymph nodes in the breast cancer data set using the semiparametric CRAFT model

<table>
<thead>
<tr>
<th>Latency: $\beta$</th>
<th>Incidence: $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5%</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.59</td>
</tr>
<tr>
<td>Trt A</td>
<td>0.16</td>
</tr>
<tr>
<td>Trt B</td>
<td>-0.46</td>
</tr>
<tr>
<td>Clinical Stage I</td>
<td>0.01</td>
</tr>
<tr>
<td>Lymph Nodes</td>
<td>-0.88</td>
</tr>
</tbody>
</table>

The results based on the above model parameters are summarized in Table 2.1. Descriptive statistics based on posterior distributions for the covariates in both the incidence and latency portions of the model include the posterior median and the posterior 2.5% and
97.5% percentiles.

Point estimates will not be compared to previous authors because the interpretation is different, as we are using the accelerated failure time model. A formal Bayesian testing for significance can be performed using Bayes factor, but conclusions may also be based on the concentration of posterior distribution around zero, namely, between the 2.5% and 97.5% percentiles. These results can be compared to p-values obtained from previous results. In general, statistical significance tends to be in agreement with those found by Peng and Dear (2000) and Lu and Ying (2004). Lu and Ying (2004) tended to produce more conservative hypothesis tests.

Figure 2.3: Breast cancer data: survival curves using (i) posterior median of the survival estimates based on the semiparametric CRAFT model (weighted by lymph nodes and clinical stage) and (ii) nonparametric Kaplan-Meier estimator, for each of the three treatments.

In terms of latency parameters, treatment A and clinical stage indication both have positive significant effects on short term survival times. Peng and Dear (2000) found these same results. For the incidence parameters based on a 0.05 cut-off criterion, both lymph nodes and clinical stage are significant with negative and positive effects; respectively, on long-term survival. Also, zero is just inside the tail cut-off value of 97.5% for the posterior distribution of treatment B, showing evidence of a positive effect on long-term survival.
These three statistics were all found to be significant or nearly significant for both Peng and Dear (2000) and Lu and Ying (2004).

In addition to descriptive statistics, the estimated survival distributions based on posterior medians are computed for each treatment, weighting by lymph nodes and clinical stage. Weighting is performed based on the corrected group prognosis method (Chang et al., 1982). For a particular treatment, four survival curves are computed corresponding to each level of lymph nodes and clinical stage. The estimated survival curve is then constructed as the weighted average of these four curves, where the weights are proportional to the number of subjects at each level of lymph nodes and clinical stage. Figure 2.3 provides this plot, overlayed with the Kaplan-Meier estimates for each treatment, illustrating that the estimated survival curves appear to be capturing the survival distribution relatively well in the long-run. Estimation in the short-run is a little off, but is still capturing the crossing survival curves feature. Figure 2.4 is a plot of the estimated error density based on posterior medians, indicating that the number and location of normal mixtures seem to be capturing the error density well.
2.8 Conclusion

We have introduced a semiparametric accelerated failure time cure model in a Bayesian framework. Modeling the error term as a mixture of normal densities provides an intuitive and useful means for semiparametric estimation. However, several mixtures densities may have been considered instead of the normal. Specifically, combining our work with that of Ghosh and Ghosal (2006), posterior consistency based on a mixture of Weibull densities holds and could have sufficed in the cure rate accelerated failure time model. Also, we assume that the same vector of covariates \( z \) is present in the incidence and latency components. In practice, it is possible to partition the covariates into these two model components (Li and Taylor, 2002) or use variable selection methods (Mitchell and Beauchamp, 1988; Kinney and Dunson, 2007) to choose appropriate subsets of covariates for the incidence and latency. The penalized prior has been adapted from Komárek et al. (2005) to control smoothing but with the smoothing parameter fixed at a reasonable level. Future work may investigate selection of an optimal smoothing mechanism and/or an optimal number of normal densities used in the mixture. And lastly, although the model has been developed for a univariate response with right-censored data, it may be extended to multivariate survival analysis or to other various censoring scenarios, such as interval-censored data.
Chapter 3

The Semiparametric AFT Model Under Interval Censoring

3.1 Introduction

Survival data commonly suffers from censoring, whereby failure times are only partially known. Estimation methods, regardless of the model chosen, need to take into account the type of censoring present in the data in order to utilize all available information and avoid biased estimates. Survival analysis methods have been extensively studied for right-censored data, while the literature on interval censoring has been more slow to follow. Partly this is because the censoring mechanism for right-censored data is much simpler than that of interval-censored data so new estimation methods are necessary. And because of the added complexities, the counting process theory that has been developed for right-censoring is generally no applicable for interval censoring (Sun, 2006). Despite the slower advances in literature for interval-censored data, this type of data is commonly found in practice. In particular, clinical trials data often involves subsequent visits to a hospital, whereby patient outcomes are only known to have occurred between two visits.

The accelerated failure time (AFT) model is a popular choice when the proportional hazards (PH) assumption is uncertain. One common complaint of the standard AFT model is its parametricity, due to its specified error distribution. Relaxing the distributional assumption has been difficult because it requires simultaneous estimation of the regression coefficients and the survival distribution. In a frequentist context, several papers have
presented various semiparametric estimation methods for the AFT model, including Rabiniwitz et al. (1995), Li and Zhang (1998), Betensky et al. (2001), Komárek et al. (2005), Lesaffre et al. (2005), Park et al. (2006), and Zhang and Davidian (2008). Penalized likelihood combining Bayesian and frequentist methods has been explored by Ghosh and Sinha (2001).

Obtaining a Bayesian solution based on posterior distribution has been extremely difficult without employing Markov chain Monte Carlo (MCMC) methods. Christensen and Johnson (1988) discuss estimation by utilizing a Dirichlet process prior (Ferguson, 1973) for the underlying baseline survival function while Johnson and Christensen (1989) points out its near impossible implementation at the time the paper was written. Since then Kuo and Mallick (1997), Walker and Mallick (1999), and Hansen and Johnson (2004) have used Dirichlet process mixtures, Polya trees, and mixtures of Dirichlet processes, respectively, as priors on the error distribution in the AFT model while using MCMC sampling methods. More recently, Komárek and Lesaffre (2007) and Komárek and Lesaffre (2008) have used finite mixtures of normal densities to estimate the error distribution for a variety of censoring scenarios. All of these papers have focused on implementation; how to obtain estimates based on the posterior distribution, while none have established theoretical justification. For right-censored data, Ghosh and Ghosal (2006) establish posterior consistency using an infinite mixture of Weibull densities under the assumption that the true underlying error distribution is in fact a mixture. Wu (2009) establish posterior consistency for the same model but dropping the assumption of a true underlying mixture density.

We propose a semiparametric AFT model for interval-censored data, where an infinite mixture of logistic densities is used to model the error density. Likelihood derivation and MCMC methods for obtaining estimates based on posterior distributions are explained. Using similar strategies as in Wu (2009) and given certain conditions, consistency of the posterior distribution is established for case 1 interval censoring. Simulations studies evaluate the performance of the model and the method is illustrated with a breast cosmesis data set.

3.2 The AFT Model with Unspecified Error Distribution

Letting $T$ be the time to occurrence of event, the accelerated failure time time model is
\[ \log T = \beta^T z + \epsilon, \]  

(3.1)

where \( z \) is a \( p \)-dimensional vector of covariates, \( \beta \) is a \( p \)-dimensional vector of unknown regression coefficients, and \( \epsilon \) represents measurement error. Notice that the model in Equation (3.1) is identifiable if \( z \) does not include an intercept term. That is, if \( \beta_1^T z + \epsilon_1 \overset{d}{=} \beta_2^T z + \epsilon_2 \) then by setting \( z = 0 \) we get \( \epsilon_1 \overset{d}{=} \epsilon_2 \) indicating that the error distributions are identical which in turn implies \( \beta_1 = \beta_2 \). We assume the covariate \( z \) is random with probability measure \( Q(z) \). For identifiability, we assume \( \int zQ(z)dz = 0 \). We assume that the error \( \epsilon \) follows an infinite mixture of specified error distributions with an unknown mixing distribution \( H(\cdot) \), where \( H(\cdot) \) is an unknown cumulative distribution function. Any of several densities could be used in the mixture to well approximate the error distribution. Chapter 2 uses normal densities while both Ghosh and Ghosal (2006) and Wu (2009) use Weibull densities. We choose to model \( \epsilon \) as a mixture of logistic distributions, since the logistic distribution has wider tails than the standard normal and its cumulative distribution function may be written in closed form for faster computational time. Then the probability density function, \( g(\epsilon) \) of \( \epsilon \) is given by

\[ g(\epsilon) \equiv g_H(\epsilon \mid \lambda) = \int \frac{e^{-\frac{(\epsilon-\mu)}{\lambda}}}{\{1 + e^{-\frac{(\epsilon-\mu)}{\lambda}}\}} z dH(\mu), \]  

(3.2)

where \( \lambda \) is the scale parameter. Notice that \( g_H(\cdot \mid \lambda) \) is identifiable in the sense that if \( g_{H_1}(\cdot \mid \lambda) = g_{H_2}(\cdot \mid \lambda) \) then \( H_1(\cdot) = H_2(\cdot) \) (Teicher, 1961). Thus the model in Equation (3.1) combined with Equation (3.2) is identifiable as a function of \( \beta, D(\cdot) \). In terms of prior specification, we assume that the mixing distribution \( H(\cdot) \sim \Pi_H \). Given \( H, \mu \sim H \). Define the precision parameter \( \lambda \sim \Pi_\lambda \) which may depend on the sample size and the regression parameter \( \beta \sim \Pi_\beta \). Let \( \Pi \) stand for \( \Pi_H \times \Pi_\lambda \times \Pi_\beta \). Later we will propose suitable priors and make additional assumptions to establish posterior consistency.

Having defined the AFT model and the error mixture, the corresponding survival function for survival times \( T = t \) is

\[ P(T > t \mid Z = z) = S(t \mid \beta, H(\cdot), \lambda, z) = \int \frac{e^{-\frac{(\log t - \mu - \beta^T z)}{\lambda}}}{1 + e^{-\frac{(\log t - \mu - \beta^T z)}{\lambda}}} dH(\mu), \]

(3.3)
3.3 Prior Distribution

Because infinite mixture distributions can often be well approximated by finite, possibly sample size-dependent, discrete distribution (Li and Barron, 2000; Komárek et al., 2005) for computation purposes we approximate the error density in Equation (3.2) by

\[ g_{L_n}(\epsilon) = \sum_{l=1}^{L_n} w_l \frac{e^{-(\epsilon - \mu_l)/\lambda}}{\left(1 + e^{-(\epsilon - \mu_l)/\lambda}\right)^2}, \]

(3.4)

where \( L_n \) is the number of normal mixtures, \( \mu_1 < \cdots < \mu_{L_n} \) is a suitably chosen ordered sequence of knots in \( \mathbb{R} \), and \( w = (w_1, \ldots, w_{L_n})^T \) are unknown non-negative mixture coefficients satisfying the restriction \( \sum_{l=1}^{L_n} w_l = 1 \). This guarantees that \( g_{L_n}(\epsilon) \) is a density function. We use the notation \( L_n \) to indicate that the number of logistic densities used in the mixture may increase with the sample size. In terms of prior specification, we use a Dirichlet distribution to model the mixture coefficients \( w \), which is an approximation to an infinite mixture of Dirichlet process priors (Ishwaran and Zarepour, 2002) which could have been used for \( \Pi_H \) had the error density not been discretized. Note that this Dirichlet prior on the mixture coefficients is an alternative to the multivariate logit distribution used in Chapter 2. We consider a noninformative prior for \( \beta \), where \( \beta \sim N(0, D_\beta) \) and the variance matrix \( D_\beta \) is adjusted to provide reasonable parameter spaces. Komárek et al. (2005)’s mixture of normals set \( \lambda = .2 \) for all analyses. Alternatively, we develop a prior for \( \lambda \) with the property that \( \lambda \) decreases with the sample size. We set \( \lambda = \kappa n^{-1/5} \) where \( \kappa \sim Uniform(a_0, b_0) \) and \( a_0 > 0 \) and \( b_0 > 0 \) are suitably chosen so \( \lambda \) is still in a general neighborhood of .2. For the rest of the paper we define the set of parameters to be estimated as \( \theta = \{\beta, \kappa, w\} \in \{\mathbb{R}^p \times (a_0, b_0) \times \mathbb{R}^{L-1}\} \).

3.4 Semiparametric Estimation Using the Posterior Distribution

We consider both case 1 and case 2 interval censoring scenarios. Suppose that each patient periodically visits a hospital, and at each patient visit, information is obtained as to whether that patient has experienced the event yet. Case 1 censoring is the case where only one such visit occurs, so we observe the time at which they visit the hospital and whether the event (i) occurred before the visit or (ii) has yet to occur. Case 2 censoring is the case
where the patient has at least two hospital visits. Here we observe these visit times, as well as whether the event (i) occurred before the first visit, (ii) occurred some time between two visits, or (iii) has yet to occur some time after the final visit.

To formalize, let $T_i$ denote the survival time of the $i$th subject and $Z_i$ a vector of covariates. For case 1 censoring, we observe the triplet $\{C_i, \Delta_i, Z_i\}$ where $C_i$ is the random censoring time and $\Delta_i = I(T_i \leq C_i)$. Given $Z = z_i$, we assume that $T_i$ is independent of $C_i$. We denote these observed data as $O_1 = \{C_i, \Delta_i, Z_i, i = 1, \ldots, n\}$ which are assumed to be independent across $i = 1, \ldots, n$.

For case 2 censoring, we observe $\{U_i, V_i, \Delta_{1i}, \Delta_{2i}, \Delta_{3i}, Z_i\}$ where $U_i$ and $V_i$ are two random censoring times with $0 < U_i < V_i < \infty$, $\Delta_{1i} = I(T_i < U_i)$, $\Delta_{2i} = I(U_i < T_i < V_i)$, and $\Delta_{3i} = I(T_i > V_i)$. Note that $\Delta_{3i}$ is redundant but is defined for clarity when specifying the likelihood function. Given $Z = z_i$, we assume that $T_i, U_i$ are $V_i$ are mutually independent. We denote these observed data as $O_2 = \{U_i, V_i, \Delta_{1i}, \Delta_{2i}, \Delta_{3i}, Z_i, i = 1, \ldots, n\}$ which are assumed to be independent across $i = 1, \ldots, n$.

Then using the finite discrete distribution for the mixing distribution, the corresponding survival function for the $i$th observation of an observed data point given the parameter $\theta$ is

$$S_{L_n}(x_i \mid \theta, z) = \sum_{i=1}^{L_n} w_i \frac{e^{-(\log x_i - \mu_1 - \beta^T z)/\lambda}}{1 + e^{-(\log x_i - \mu_1 - \beta^T z)/\lambda}}.$$ 

(3.5)

Hence the observed likelihoods based on case 1 and case 2 interval-censored data are

$$L(\theta \mid O_1) = \prod_{i=1}^{n} \left\{ F_{L_n}(c_i \mid \theta, z) \right\}^{\delta_1} \left\{ S_{L_n}(c_i \mid \theta, z) \right\}^{1-\delta_1},$$

(3.6)

$$L(\theta \mid O_2) = \prod_{i=1}^{n} \left\{ F_{L_n}(u_i \mid \theta, z) \right\}^{\delta_1} \left\{ F_{L_n}(v_i \mid \theta, z) - F_{L_n}(u_i \mid \theta, z) \right\}^{\delta_2} \left\{ S_{L_n}(v_i \mid \theta, z) \right\}^{\delta_3},$$

(3.7)

where $S_{L_n}(x_i \mid \theta, z)$ is defined above in Equation (3.5) and $F_{L_n}(x_i \mid \theta, z) = 1 - S_{L_n}(x_i \mid \theta, z)$.

Having specified the observed likelihoods in Equations (3.6) and (3.7) and prior distributions in the previous section, the posterior distribution of the parameters $\theta$ for either case 1 or case 2 interval-censored data is

$$\pi(\theta \mid O_j) \propto L(\theta \mid O_j)\pi(\beta)\pi(\kappa)\pi(w) \quad \text{for } j = 1, 2.$$
As in Chapter 2, the posterior distribution of the parameters can not be obtained in closed form. Using the finite discrete approximation to the infinite logistic mixture allows WinBUGS to perform Markov chain Monte Carlo sampling from the posterior distribution. As shown explicitly in Ghosh and Ghosal (2006), latent variables $L^* = (L^*_1, \ldots, L^*_n)$ which indicate the group membership to the $\mu_l$ node along with probability vector $w = (w_1, \ldots, w_{L^n})^T$ are introduced. Note that $\text{pr}(L^*_l = l) = w_l$ for $l = 1, \ldots, L_n$. This hierarchical format is easier to implement in WinBUGS.

We select knots ranging from $\mu_1 = -M$ to $\mu_{L_n} = M$ where $\mu_l = -M + 2M(l - 1)/(L_n - 1), l = 1, \ldots, L_n$ for some suitably chosen large $M > 0$. This should be a wide enough range to account for densities that may have large tails, such as the extreme value or logistic distributions. Ishwaran and Zarepour (2002) suggest using $L_n = n^{1/2}$ for large $n$ and $L_n = n$ for small $n$ while Ghosh and Ghosal (2006) say that more work is necessary to determine an optimal number. We will select an order of $L_n$ in a similar fashion, that should allow each logistic density to overlap with a few of its neighborhoods and may increase with $n$.

3.5 Consistency of Posterior Distribution for Case 1 Interval Censoring

As discussed in Section 2.5, posterior consistency is a desirable large sample property of the posterior distribution, guaranteeing that the posterior distribution will concentrate in an arbitrarily small neighborhood of the true value of the parameter. Hence, with a large amount of data, the posterior distribution is nearing the truth so accurate estimation is attainable. Section 2.5 provides a formal definition of posterior consistency. Sufficient conditions for achieving posterior consistency involving appropriate tests and the prior positivity of a neighborhood defined by the Kullback-Leibler divergence have been used extensively since their introduction by Schwartz (1965).

Nice discussions and examples surrounding posterior consistency can be found in Diaconis and Freedman (1986), Barron et al. (1999), and Ghosal (2000). Applications of posterior consistency to Bayesian density estimation in mixture models has been studied by Barron et al. (1999), Ghosal et al. (1999a), Ghosh and Ramamoorthi (2003), Amewou-Atisso et al. (2003), and Wu and Ghosal (2008). Consistency in survival models without
covariates has been studied by Ghosh et al. (1994), Ghosal et al. (1999b), and Kim and Lee (2001) while consistency for survival models with covariates has been studied by Ghosh and Ghosal (2006) and Wu (2009). In particular, with varying assumptions, these latter two have provided proofs of posterior consistency for the AFT model with an infinite mixture of Weibull distributions for right-censored data. Because Wu (2009)’s proof does not require the assumption that the true underlying error distribution is indeed a mixture itself, we use some similar arguments to establish both the existence of appropriate tests and the Kullback-Leibler property in order to verify posterior consistency for an AFT model with case 1 interval-censored data and an infinite mixture of logistic distributions.

For notation, let $F_c$ and $f_c$ denote the cumulative distribution and density functions of the censoring time $C$. Let $S_0$ and $F_0$ be the corresponding true survival and cumulative distribution functions for $S$ and $F$, where $S$ is defined in Equation (3.3) and $F(\cdot) = 1 - S(\cdot)$. Then the following main theorem verifies that the posterior distribution is consistent.

**Theorem 2.** Consider the AFT model with $\Pi$ as defined in Section 3.2 with infinite mixture of logistic densities and case 1 interval-censored data. Suppose

1. the covariate $Z$ is compactly supported, and for any quadrant $\Gamma$, $Q(\Gamma\{\|z\| < \zeta\}) > 0$ for some $\zeta > 0$;
2. for some $0 < M < \infty$, $0 < f_0(x) < M$ for all $x$;
3. $\int f_c(x + \xi)\log S_0(x)dx < \infty$, and $\int f_c(x + \xi)\log F_0(x)dx < \infty$ for all $\xi \in \mathbb{R}$;
4. $\int |x|dH(x) < \infty$ almost surely $\Pi$;
5. the weak support of $\Pi$ is the space of all probability measures on $\mathbb{R}$;
6. for some $\delta > 0$ and any $\xi \in \mathbb{R}$, $\int f_c(x + \xi)\log \frac{S_{0,0}(x)}{\phi_{0}(x)}dx < \infty$ and $\int f_c(x + \xi)\log \frac{F_{0,0}(x)}{\phi_{0}(x)}dx < \infty$, where $\phi_{0}(x) := \inf_{[s-x, s+\delta \lambda_m]} f_0(s)$;
7. $\int |x|^{\eta+1}f_c(x + \xi)dx < \infty$ for all $\xi \in \mathbb{R}$ and some $\eta > 0$;

Then for any weak neighborhood $\mathcal{U}$ of $F_0$,

$$
\Pi\{(F, \beta) : F \in \mathcal{U}, \|\beta - \beta_0\| < \epsilon|(X_i, Z_i, \Delta_i), \ldots, (X_n, Z_n, \Delta_n)\} \to 1
$$

almost surely in $P_{\hat{f}_0(z, x), f_c}$-probability
A note on Theorem 2. We assume that the domains of $Z, \beta$ and $\lambda$, and the support of $H$ are compact. Also, we assume that the vector $Z$ with zero as one of its components is a possible value of the covariate $Z$, which if not, the covariates may be shifted to satisfy this condition (conditions 1 and 5). The failure times $x \in \mathbb{R}$ must satisfy $\int |x|dH(x) < \infty$ almost surely $\Pi$ (condition 4). We assume the true underlying error density, $f_0$ is sufficiently bounded by some constant $M$ (condition 2). And lastly we assume $S_0, F_0, f_0,$ and $f_c$ must satisfy conditions 3, 6, and 7. The proof of Theorem 2 is given in Appendix C.1.

### 3.6 Simulation Studies

We conduct several simulations to analyze how well the estimators obtained from the posterior distributions are performing under a variety of settings. We consider two covariates, $z_1$ is uniformly distributed between 0 and 1 and $z_2$ is a 0-1 binary variable from a Bernoulli distribution with mean 0.5. Covariates are centered by their population mean, which is 0.5 for both variables. Such centering is known to reduce the correlations between the regression coefficients and hence, leads to more efficient Markov chain Monte Carlo sampling schemes (Roberts and Sahu, 2001).

The AFT model is $\log T = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \epsilon$ where $\beta_0 = -1.0, \beta_1 = -1.0$ and $\beta_2 = 1.0$. We consider normal (symmetric), logistic (symmetric), extreme value (skewed), and a mixture of two normal (bimodal) distributions for $\epsilon$, each of which has mean 0 and variance 0.25. More specifically, these include (i) Normal(0,0.25) with density function $g(\epsilon; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{(\epsilon - \mu)^2/(2\sigma^2)\}$, (ii) Logistic(0, 0.28) with density function $g(\epsilon; a, b) = \exp[-(\epsilon - a)/b]/(b[1 + \exp\{-(\epsilon - a)/b\}]^2)$, (iii) EV(0.23, 0.39) where EV represents the extreme value distribution with density function $g(\epsilon; a, b) = b^{-1} \exp\{(\epsilon - a)/b\} \exp[-\exp\{(\epsilon - a)/b\}]$, and (iv) mixture of two normal densities: $N(-0.8, 0.09)$ and $N(0.2, 0.09)$, with density function $g(\epsilon; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = 0.2(2\pi\sigma_1^2)^{-1/2} \exp\{(\epsilon - \mu_1)^2/(2\sigma_1^2)\} + 0.8(2\pi\sigma_2^2)^{-1/2} \exp\{(\epsilon - \mu_2)^2/(2\sigma_2^2)\}$. The bimodal distribution is weighted disproportionately in order to form a density similar to the underlying density estimated in the breast cosmesis data set discussed in the next chapter.

For case 1 censoring scenarios, the single censoring time is generated from a uniform(0,0.85) distribution. This results in roughly 50% left and 50% right censoring percentages. For case 2 censoring, the smallest time is generated from a uniform(0,0.55)
distribution while the larger time is generated from the sum of the first generated time point and a uniform(0,0.65) distribution. This results in roughly 33% left, 33% interval, and 33% right-censoring percentages. We use sample sizes of $n = 200$ or 400 subjects. In terms of prior specification, we set $\beta_1$ and $\beta_2$, $\sim N(0,3)$. Also, $a_0 = 0.65$ and $b_0 = 0.75$ so $\kappa \sim Uniform(0.65,0.75)$ for case 1 interval censoring while for case 2 interval censoring, $a_0 = 0.60$ and $b_0 = 0.80$ so $\kappa \sim Uniform(0.60,0.80)$. Recall $\lambda = \kappa n^{-1/5}$. The mixture of normal densities used to estimate the error term range from $\mu_l = -4,\ldots,4$ with $L_n = 17$ nodes.

For case 1 simulations, the posterior generation in WinBUGS is based on 8000 burn-ins and an additional 4000 runs from three parallel chains, for a total of 12,000 Markov chain Monte Carlo samples per simulated data. For case 2 simulations, the posterior generation in WinBUGS is based on 4000 burn-ins and an additional 2000 runs from three parallel chains, for a total of 6,000 Markov chain Monte Carlo samples per simulated data.

For all these simulation scenarios, 100 Monte Carlo samples were generated to study the sampling variability of the posterior estimates of $\beta_1$, and $\beta_2$. In order to measure the empirical performance of these posterior estimates, we computed the bias of the posterior median, the Monte Carlo standard error of the posterior median, the Monte Carlo average of the posterior standard deviation, and the nominal coverage probabilities of covering the true value based on the 95% posterior interval.

Tables C.1-C.2 in Appendix C.3 provide results to the Monte Carlo simulations. The posterior medians perform well in all simulations. For the case 1 censoring cases, estimates tend to be slightly biased in the $n = 200$ case but disappear when $n = 400$. Overall, the posterior medians perform well and any biases are due to small sample sizes. Coverage probabilities are close to their nominal values in most settings. Any smaller probabilities could possibly be due to Monte Carlo error. The Monte Carlo standard errors of the posterior medians and Monte Carlo averages of the posterior standard deviations are fairly close and decrease for the large sample size. Also, the estimated mixture densities capture the true underlying error distributions quite well for all simulation scenarios. As illustration, Figure 3.1 shows the estimated and true survival distributions for the logistic, extreme value, and bimodal error densities when $n = 400$ for case 2 interval censoring.
3.7 Breast Cosmesis Data Set

A breast cosmesis data set with interval-censored observations originally presented by Finkelstein and Wolfe (1985) is analysed to demonstrate the proposed semiparametric AFT model. Since its introduction, many papers have re-analysed the data set using several methods, including the PH model (Finkelstein and Wolfe, 1985; Goggins et al., 1998; Goetghebeur and Ryan, 2000; Betensky et al., 2002), AFT model (Tian and Cai, 2006), Bayesian discretized semiparametric models (Sinha et al., 1999), and more recently by a smoothed regression approach which includes the PH, AFT, and proportional odds models (Zhang and Davidian, 2008).

A total of 94 early breast cancer patients are included in this study, where time to moderate or severe breast retraction (measure of cosmetic deterioration) is used as the failure endpoint. Two treatments are compared, resulting in a single covariate for the analysis. There are 46 women receiving radiation treatment ($z=0$) and 48 women receiving radiation combined with adjuvant chemotherapy ($z=1$). Observations are recorded every four to six weeks, with intervals tending to increase over time. At each visit, an indicator of whether a patient has experienced cosmetic deterioration since the previous visit is recorded. There were 56 women who experienced cosmetic deterioration, resulting in left or interval-censored observations, while 38 women never experienced the event of interest and were right-censored.
We fit the breast cosmesis data set using the semiparametric AFT model. The logistic mixture nodes range from \(\min\{\log(u_i)I(\delta_3_i = 0)\} - 2\) to \(\max\{\log(v_i)I(\delta_1_i = 0)\} + 2\), which should allow the mixture distribution to easily capture the spread of the data. Varying \(L_n\) to equal 17 or 25 gave nearly identical results, so we present the results when \(L_n = 17\). In terms of prior specification, we set \(\beta \sim N(0,3)\). Also, \(\lambda = \kappa n^{-1/5}\) where \(\kappa \sim \text{Uniform}(0.6,0.8)\) and the mixture coefficients \(w \sim \text{Dirichlet}(2/17,\ldots,2/17)\). For each simulation, the posterior generation in WinBUGS is based on 50,000 burn-ins and an additional 50,000 runs from three parallel chains, for a total of 150,000 Markov chain Monte Carlo samples per simulated data.

Several diagnostic measures are examined to verify the fit of the model and can be found in Appendix C.4. Using the CODA package available in R, the Gelman-Rubin shrink factors for all statistics are \(\leq 1.01\), indicating good mixing and convergence to the appropriate model. Figure C.1 provides the trace and posterior density plots for \(\beta\), which indicate that the chains have converged. Figure C.2 provides a plot of the posterior means of the mixture coefficients, suggesting that the location of the logistic nodes seem to be capturing the data.

The results for this simulation are summarized in Table 3.1. Descriptive statistics based on the posterior distribution of treatment include the posterior median and the posterior 2.5% and 97.5% percentiles. The point estimate tends to be very much inline with Tian and Cai (2006) who also fit an AFT model. In addition to summary statistics, we compute the posterior median survival estimates. Figure 3.2 displays these estimated survival curves for each of the two treatments as well as Turnbull’s nonparametric estimates. The crossing survival curves shown with the nonparametric estimates can not be captured with an AFT model. However, our survival estimates are still able to pick up the general survival trend. The number and location of logistic density nodes appear to be capturing the error density well, as shown by a plot in Figure 3.3 of the posterior median error density estimates.
Figure 3.2: Interval-censored breast cosmesis data: survival curves using (i) posterior median of the survival estimates based on the semiparametric AFT model and (ii) Turnbull nonparametric estimator, for each of the two treatments.

Figure 3.3: Interval-censored breast cosmesis data: posterior median error density estimates from the semiparametric AFT model using a mixture of 17 logistic distributions
3.8 Conclusion

Mixture density estimation is an intuitive and effective way to implement semiparametric estimation, and also useful alternative to parametric estimation which is commonly associated with the AFT model. Although the logistic density was used and posterior consistency verified, many continuous densities may have possibly sufficed in the mixture. In particular, Wu (2009)'s proof of posterior consistency using a mixture of Weibull densities could possibly be extended to the interval censoring case. Future work could extend the proof to the more general case 2 interval-censored data, or verify consistency for a class of densities rather than merely the logistic. Lastly, the model could be extended to multivariate survival analysis, doubly censored data, or re-introducing a cure fraction now with interval-censored data.
Bibliography


Appendices
Appendix A

Supplement to Chapter 1

A.1 Derivation of Standard Errors based on the EM algorithm

We present Oakes (1999)’s method for computing the information matrix when using the EM algorithm. Define the observed data $z$, complete data $y$, and observed data log-likelihood $l(\gamma, \beta; F, z)$. Define the parameter space $\theta = \{\gamma, \beta\}$. For notation purposes, we shorten the following quantities:

$$
p_i = p(\gamma, z_i) = \frac{\exp(\gamma'z_i)}{1 + \exp(\gamma'z)}$$
$$
\lambda_i = \lambda(\log t_i - \beta'z_i)$$
$$
S_i = S(\log t_i - \beta'z_i)
$$

The EM algorithm maximizes the ‘Q function’:

$$
Q(\theta' | \theta) = E_{y|z} l_c(\theta'; x) = l_{c_1}(\gamma') + l_{c_2}(\beta')
$$

(A.1)

where

$$
l_{c_1}(\gamma') = \sum_{i=1}^{n} w_i^{(m)} \log(p'_i) + (1 - w_i^{(m)}) \log(1 - p'_i),
$$

(A.2)

$$
l_{c_2}(\beta') = \sum_{i=1}^{n} \delta_i \log(\lambda_i) + w_i^{(m)} \log(S_i),
$$

(A.3)

$$
w_i^{(m)} = \delta_i + (1 - \delta_i) \frac{p_i S_i}{1 - p_i + p_i S_i}_{(\theta^{(m)}, z)}
$$

(A.4)

Each step of the EM algorithm updates $\gamma'$ and $\beta'$ while holding $w_i^{(m)}$ fixed, where $w_i^{(m)}$ is a function of the current estimates of $\gamma$ and $\beta$. 
The information matrix derived in Oakes (1999) is
\[
\frac{\partial^2 L(\theta, z)}{\partial \theta^2} = \left\{ \frac{\partial^2 Q(\theta'|\theta)}{\partial \theta'^2} + \frac{\partial^2 Q(\theta'|\theta)}{\partial \theta' \partial \theta} \right\}_{\theta' = \theta} \tag{A.5}
\]

Let \( \beta_j, \gamma_j \) and \( z_{ji} \) denote the \( j \)th parameters and covariates where \( j = 0, 1, 2 \) and \( z_0 = 1 \). We provide the necessary first and second partial derivatives to insert into Equation (A.5). The first derivatives are as follows:

\[
\frac{\partial Q}{\partial \beta'_j} = -\delta_i \dot{\lambda}_i z_{ji} + w_i^{(m)} \lambda_i \dot{z}_{ji} \quad \frac{\partial Q}{\partial \gamma'_j} = (w_i^{(m)} - p_i) z_{ji}
\]

For the first term in Equation (A.5), we calculate:

\[
\frac{\partial Q^2}{\partial \beta'_j \partial \beta'_k} = -\delta_i \dot{\lambda}_i z_{ji} \quad \frac{\partial Q^2}{\partial \gamma'_j \partial \gamma'_k} = -p_i (1 - p_i) z_{ki} z_{ji} \quad \frac{\partial Q^2}{\partial \gamma'_j \partial \beta'_k} = 0
\]

For the second term in Equation (A.5), we calculate:

\[
\frac{\partial w_i^{(m)}}{\partial \beta_j} = (1 - \delta_i) \left[ \frac{-(1 - p_i + p_i S_i)p_i \dot{S}_i z_{ji} + p_i^2 S_i \dot{S}_i z_{ji}}{(1 - p_i + p_i S_i)^2} \right] \quad \frac{\partial w_i^{(m)}}{\partial \gamma_j} = \frac{(1 - \delta_i)p_i(1 - p_i)S_i z_{ji}}{(1 - p_i + p_i S_i)^2}
\]

\[
\frac{\partial Q^2}{\partial \beta'_j \partial \beta_k} = \frac{\partial w_i^{(m)}}{\partial \beta_k} \lambda_i z_{ji} \quad \frac{\partial Q^2}{\partial \beta'_j \partial \gamma_k} = \frac{\partial w_i^{(m)}}{\partial \gamma_k} \lambda_i z_{ji} \quad \frac{\partial Q^2}{\partial \gamma'_j \partial \beta_k} = \frac{\partial w_i^{(m)}}{\partial \beta_k} z_{ji} \quad \frac{\partial Q^2}{\partial \gamma'_j \partial \gamma_k} = \frac{\partial w_i^{(m)}}{\partial \gamma_k} z_{ji}
\]

The second order partial derivatives above are redundant, because the second term in Equation (A.5) is a symmetric matrix. Additionally, \( \dot{S}_i \) is the first derivative of the survival function and \( \dot{\lambda}_i \) and \( \ddot{\lambda}_i \) are the first and second derivatives of the hazard function \( \lambda_i \), which are dependent on the density of the errors. The chain rule involving the covariates has already been taken account of and has been inserted above accordingly. These derivatives are:

\[
\dot{S}_i = -f_i \quad \dot{\lambda}_i = \frac{S_i \dot{f}_i + f_i^2}{S_i^2} \quad \ddot{\lambda}_i = \frac{S_i \ddot{f}_i + \dot{f}_i f_i}{S_i^2} + 2 \lambda_i \dot{\lambda}_i
\]

See appendix A.2 for specific error densities and the associated derivatives.
A.2 Densities and Derivatives of the Error Distribution

Table A.1: Densities and their respective first and second derivatives for the error distribution

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Density Function</th>
<th>First Derivative</th>
<th>Second Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Normal</td>
<td>$g(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$</td>
<td>$\dot{g}(x) = -\frac{xe^{-x^2/2}}{\sqrt{2\pi}}$</td>
<td>$\ddot{g}(x) = -\frac{e^{-x^2/2}(x^2 - 1)}{\sqrt{2\pi}}$</td>
</tr>
<tr>
<td>Extreme Value(0,1)</td>
<td>$g(x) = e^x e^{-e^x}$</td>
<td>$\dot{g}(x) = e^x e^{-e^x} (1 - e^x)$</td>
<td>$\ddot{g}(x) = e^x e^{-e^x} (1 - 3e^x + e^{2x})$</td>
</tr>
<tr>
<td>Logistic(0,1)</td>
<td>$g(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$</td>
<td>$\dot{g}(x) = \frac{-e^{-x} + e^{-2x}}{(1 + e^{-x})^3}$</td>
<td>$\ddot{g}(x) = \frac{e^{-x} (1 - 4e^{-x} + e^{-2x})}{(1 + e^{-x})^4}$</td>
</tr>
</tbody>
</table>
### A.3 Simulation Results for Frequentist Estimation Methods

Table A.2: Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the standard normal distribution.

#### ML Estimation

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>0.006</td>
<td>0.287</td>
<td>0.286</td>
<td>1.071</td>
<td>0.948</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>0.005</td>
<td>0.467</td>
<td>0.450</td>
<td>1.208</td>
<td>0.926</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>0.004</td>
<td>0.273</td>
<td>0.263</td>
<td>1.083</td>
<td>0.948</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.099</td>
<td>0.619</td>
<td>0.596</td>
<td>1.591</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>0.011</td>
<td>0.945</td>
<td>0.929</td>
<td>1.916</td>
<td>0.958</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.098</td>
<td>0.550</td>
<td>0.543</td>
<td>1.508</td>
<td>0.964</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.011</td>
<td>0.319</td>
<td>0.311</td>
<td>1.080</td>
<td>0.952</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>0.005</td>
<td>0.517</td>
<td>0.490</td>
<td>1.258</td>
<td>0.938</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.002</td>
<td>0.305</td>
<td>0.289</td>
<td>1.096</td>
<td>0.938</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.056</td>
<td>0.531</td>
<td>0.526</td>
<td>0.592</td>
<td>0.960</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.018</td>
<td>0.898</td>
<td>0.840</td>
<td>1.770</td>
<td>0.948</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.042</td>
<td>0.473</td>
<td>0.480</td>
<td>1.309</td>
<td>0.966</td>
</tr>
</tbody>
</table>

#### EM Estimation

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>0.006</td>
<td>0.287</td>
<td>0.285</td>
<td>1.071</td>
<td>0.976</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>0.005</td>
<td>0.467</td>
<td>0.448</td>
<td>1.208</td>
<td>0.926</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>0.004</td>
<td>0.273</td>
<td>0.263</td>
<td>1.083</td>
<td>0.948</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.099</td>
<td>0.618</td>
<td>0.589</td>
<td>1.591</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>0.012</td>
<td>0.945</td>
<td>0.924</td>
<td>1.915</td>
<td>0.958</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.098</td>
<td>0.549</td>
<td>0.539</td>
<td>1.508</td>
<td>0.964</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.011</td>
<td>0.319</td>
<td>0.310</td>
<td>1.079</td>
<td>0.978</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>0.004</td>
<td>0.517</td>
<td>0.488</td>
<td>1.258</td>
<td>0.938</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.002</td>
<td>0.305</td>
<td>0.288</td>
<td>1.096</td>
<td>0.938</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.056</td>
<td>0.531</td>
<td>0.523</td>
<td>0.592</td>
<td>0.960</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.018</td>
<td>0.898</td>
<td>0.838</td>
<td>1.770</td>
<td>0.948</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.041</td>
<td>0.473</td>
<td>0.479</td>
<td>1.309</td>
<td>0.966</td>
</tr>
</tbody>
</table>
Table A.3: Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the extreme value distribution.

### ML Estimation

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 1.0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>-0.003</td>
<td>0.294</td>
<td>0.292</td>
<td>1.093</td>
<td>0.950</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>-0.021</td>
<td>0.452</td>
<td>0.458</td>
<td>1.246</td>
<td>0.950</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>-0.005</td>
<td>0.249</td>
<td>0.263</td>
<td>1.052</td>
<td>0.970</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.045</td>
<td>0.586</td>
<td>0.566</td>
<td>1.435</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>0.038</td>
<td>0.918</td>
<td>0.883</td>
<td>1.919</td>
<td>0.954</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.043</td>
<td>0.544</td>
<td>0.516</td>
<td>1.384</td>
<td>0.944</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>-0.017</td>
<td>0.323</td>
<td>0.319</td>
<td>1.139</td>
<td>0.948</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>-0.002</td>
<td>0.511</td>
<td>0.498</td>
<td>1.264</td>
<td>0.948</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>-0.031</td>
<td>0.297</td>
<td>0.290</td>
<td>1.027</td>
<td>0.942</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.003</td>
<td>0.527</td>
<td>0.507</td>
<td>0.530</td>
<td>0.942</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>0.030</td>
<td>0.842</td>
<td>0.807</td>
<td>1.769</td>
<td>0.946</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.020</td>
<td>0.456</td>
<td>0.461</td>
<td>1.248</td>
<td>0.964</td>
</tr>
</tbody>
</table>

### EM Estimation

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 1.0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>-0.003</td>
<td>0.294</td>
<td>0.290</td>
<td>1.093</td>
<td>0.978</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>-0.021</td>
<td>0.452</td>
<td>0.456</td>
<td>1.246</td>
<td>0.950</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>-0.005</td>
<td>0.249</td>
<td>0.263</td>
<td>1.052</td>
<td>0.970</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.045</td>
<td>0.586</td>
<td>0.561</td>
<td>1.435</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>0.038</td>
<td>0.917</td>
<td>0.880</td>
<td>1.919</td>
<td>0.954</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.043</td>
<td>0.544</td>
<td>0.513</td>
<td>1.384</td>
<td>0.944</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>-0.017</td>
<td>0.323</td>
<td>0.316</td>
<td>1.138</td>
<td>0.976</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>-0.002</td>
<td>0.511</td>
<td>0.495</td>
<td>1.264</td>
<td>0.948</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>-0.031</td>
<td>0.297</td>
<td>0.289</td>
<td>1.027</td>
<td>0.942</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.003</td>
<td>0.527</td>
<td>0.506</td>
<td>0.530</td>
<td>0.942</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>0.030</td>
<td>0.842</td>
<td>0.805</td>
<td>1.769</td>
<td>0.946</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.020</td>
<td>0.456</td>
<td>0.460</td>
<td>1.248</td>
<td>0.964</td>
</tr>
</tbody>
</table>
Table A.4: Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when \( n = 100 \) and \( \epsilon \) is generated from the logistic(0,1) distribution

### ML Estimation

<table>
<thead>
<tr>
<th>( \gamma_0 = 1 )</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>0.005</td>
<td>0.557</td>
<td>0.549</td>
<td>1.300</td>
<td>0.944</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>0.009</td>
<td>0.906</td>
<td>0.858</td>
<td>1.801</td>
<td>0.936</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>-0.004</td>
<td>0.474</td>
<td>0.498</td>
<td>1.217</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.043</td>
<td>0.722</td>
<td>0.761</td>
<td>1.609</td>
<td>0.972</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>0.135</td>
<td>1.171</td>
<td>1.196</td>
<td>2.659</td>
<td>0.972</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.035</td>
<td>0.693</td>
<td>0.706</td>
<td>1.551</td>
<td>0.964</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \gamma_0 = 0.5 )</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.011</td>
<td>0.623</td>
<td>0.598</td>
<td>1.366</td>
<td>0.948</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>0.000</td>
<td>1.054</td>
<td>0.930</td>
<td>2.112</td>
<td>0.908</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>-0.015</td>
<td>0.570</td>
<td>0.547</td>
<td>1.296</td>
<td>0.940</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.025</td>
<td>0.660</td>
<td>0.635</td>
<td>0.711</td>
<td>0.950</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>0.057</td>
<td>1.047</td>
<td>1.007</td>
<td>2.214</td>
<td>0.968</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.003</td>
<td>0.561</td>
<td>0.573</td>
<td>1.320</td>
<td>0.964</td>
</tr>
</tbody>
</table>

### EM Estimation

<table>
<thead>
<tr>
<th>( \gamma_0 = 1 )</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>0.004</td>
<td>0.555</td>
<td>0.543</td>
<td>1.301</td>
<td>0.984</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>0.011</td>
<td>0.901</td>
<td>0.851</td>
<td>1.789</td>
<td>0.936</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>-0.006</td>
<td>0.471</td>
<td>0.496</td>
<td>1.210</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.040</td>
<td>0.716</td>
<td>0.725</td>
<td>1.594</td>
<td>0.970</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>0.136</td>
<td>1.158</td>
<td>1.148</td>
<td>2.632</td>
<td>0.972</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.039</td>
<td>0.680</td>
<td>0.669</td>
<td>1.543</td>
<td>0.964</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \gamma_0 = 0.5 )</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.011</td>
<td>0.621</td>
<td>0.593</td>
<td>1.364</td>
<td>0.984</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>0.001</td>
<td>1.048</td>
<td>0.923</td>
<td>2.098</td>
<td>0.906</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>-0.016</td>
<td>0.567</td>
<td>0.544</td>
<td>1.290</td>
<td>0.944</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.023</td>
<td>0.655</td>
<td>0.618</td>
<td>0.703</td>
<td>0.950</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>0.058</td>
<td>1.036</td>
<td>0.986</td>
<td>2.193</td>
<td>0.968</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.004</td>
<td>0.556</td>
<td>0.562</td>
<td>1.318</td>
<td>0.964</td>
</tr>
</tbody>
</table>
Table A.5: Simulation results for the parametric CRAFT model based on ML and EM-based estimation with 500 Monte Carlo simulations when $n = 100$ and $\epsilon$ is generated from the logistic(0,0.7) distribution

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML Estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_0 = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>$-1$</td>
<td>0.018</td>
<td>0.360</td>
<td>0.354</td>
<td>1.095</td>
<td>0.946</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$-1$</td>
<td>$-0.015$</td>
<td>0.574</td>
<td>0.554</td>
<td>1.360</td>
<td>0.942</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>$-0.005$</td>
<td>0.302</td>
<td>0.321</td>
<td>1.082</td>
<td>0.972</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1</td>
<td>0.051</td>
<td>0.635</td>
<td>0.634</td>
<td>1.508</td>
<td>0.972</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>0.072</td>
<td>0.979</td>
<td>0.980</td>
<td>2.108</td>
<td>0.964</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>$-1$</td>
<td>$-0.044$</td>
<td>0.601</td>
<td>0.579</td>
<td>1.451</td>
<td>0.960</td>
</tr>
<tr>
<td>$\gamma_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>$-1.0$</td>
<td>0.013</td>
<td>0.399</td>
<td>0.387</td>
<td>1.133</td>
<td>0.946</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$-1.0$</td>
<td>$-0.008$</td>
<td>0.661</td>
<td>0.603</td>
<td>1.453</td>
<td>0.918</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.0</td>
<td>$-0.025$</td>
<td>0.370</td>
<td>0.352</td>
<td>1.087</td>
<td>0.946</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.5</td>
<td>0.004</td>
<td>0.562</td>
<td>0.548</td>
<td>0.570</td>
<td>0.954</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.0</td>
<td>0.049</td>
<td>0.902</td>
<td>0.871</td>
<td>1.914</td>
<td>0.964</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>$-1.0$</td>
<td>$-0.017$</td>
<td>0.490</td>
<td>0.496</td>
<td>1.274</td>
<td>0.958</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>MSE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EM Estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_0 = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>$-1$</td>
<td>0.018</td>
<td>0.360</td>
<td>0.352</td>
<td>1.095</td>
<td>0.976</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$-1$</td>
<td>$-0.015$</td>
<td>0.574</td>
<td>0.551</td>
<td>1.359</td>
<td>0.942</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>$-0.005$</td>
<td>0.302</td>
<td>0.320</td>
<td>1.082</td>
<td>0.972</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1</td>
<td>0.050</td>
<td>0.634</td>
<td>0.622</td>
<td>1.506</td>
<td>0.972</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>0.072</td>
<td>0.978</td>
<td>0.972</td>
<td>2.107</td>
<td>0.964</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>$-1$</td>
<td>$-0.044$</td>
<td>0.600</td>
<td>0.569</td>
<td>1.450</td>
<td>0.960</td>
</tr>
<tr>
<td>$\gamma_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>$-1.0$</td>
<td>0.013</td>
<td>0.399</td>
<td>0.385</td>
<td>1.133</td>
<td>0.980</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$-1.0$</td>
<td>$-0.008$</td>
<td>0.660</td>
<td>0.600</td>
<td>1.452</td>
<td>0.918</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.0</td>
<td>$-0.025$</td>
<td>0.369</td>
<td>0.351</td>
<td>1.087</td>
<td>0.946</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.5</td>
<td>0.004</td>
<td>0.562</td>
<td>0.544</td>
<td>0.569</td>
<td>0.954</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.0</td>
<td>0.049</td>
<td>0.901</td>
<td>0.867</td>
<td>1.913</td>
<td>0.964</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>$-1.0$</td>
<td>$-0.017$</td>
<td>0.490</td>
<td>0.495</td>
<td>1.274</td>
<td>0.958</td>
</tr>
</tbody>
</table>
A.4 Simulation results for Bayesian Estimation Methods

Table A.6: Simulation results for the parametric CRAFT model based on Bayesian estimation with 500 Monte Carlo simulations when \( n = 100, \gamma_0 = 1.0 \) in the data generation, and we set the \( \gamma_0 \) prior to be \( \gamma_0 \sim N(0,10) \).

\[ \epsilon \sim N(0,1), \gamma_0 = 1.0 \]

<table>
<thead>
<tr>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>0.023</td>
<td>0.144</td>
<td>0.145</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>0.029</td>
<td>0.459</td>
<td>0.468</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>0.035</td>
<td>0.253</td>
<td>0.279</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.115</td>
<td>0.317</td>
<td>0.309</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>-0.096</td>
<td>0.992</td>
<td>0.948</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>-0.048</td>
<td>0.584</td>
<td>0.585</td>
</tr>
</tbody>
</table>

\[ \epsilon \sim \text{Logistic}(0,1), \gamma_0 = 1.0 \]

<table>
<thead>
<tr>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-1</td>
<td>0.114</td>
<td>0.351</td>
<td>0.322</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1</td>
<td>-0.061</td>
<td>0.849</td>
<td>0.849</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>0.081</td>
<td>0.561</td>
<td>0.538</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1</td>
<td>0.327</td>
<td>0.515</td>
<td>0.705</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1</td>
<td>-0.005</td>
<td>1.072</td>
<td>1.344</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1</td>
<td>0.090</td>
<td>0.786</td>
<td>1.074</td>
</tr>
</tbody>
</table>
Table A.7: Simulation results for the parametric CRAFT model based on Bayesian estimation with 500 Monte Carlo simulations when $n = 100$ or $200$, $\epsilon \sim N(0, 1)$ or logistic$(0, 1)$, $\gamma_0 = 1.0$ in the data generation, and we set the $\gamma_0$ prior to be $\gamma_0 \sim N(0, 10)$.

$\epsilon \sim N(0, 1)$, $\gamma_0 = 1.0$

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1</td>
<td>0.021</td>
<td>0.143</td>
<td>0.143</td>
<td>0.942</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1</td>
<td>0.032</td>
<td>0.457</td>
<td>0.467</td>
<td>0.958</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>0.033</td>
<td>0.251</td>
<td>0.278</td>
<td>0.960</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1</td>
<td>0.094</td>
<td>0.300</td>
<td>0.293</td>
<td>0.922</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>-0.097</td>
<td>0.980</td>
<td>0.936</td>
<td>0.934</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-1</td>
<td>-0.044</td>
<td>0.570</td>
<td>0.569</td>
<td>0.954</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1</td>
<td>0.010</td>
<td>0.099</td>
<td>0.096</td>
<td>0.934</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1</td>
<td>-0.019</td>
<td>0.306</td>
<td>0.319</td>
<td>0.952</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>0.013</td>
<td>0.184</td>
<td>0.188</td>
<td>0.948</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1</td>
<td>0.032</td>
<td>0.197</td>
<td>0.187</td>
<td>0.920</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>-0.062</td>
<td>0.601</td>
<td>0.632</td>
<td>0.968</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-1</td>
<td>0.000</td>
<td>0.374</td>
<td>0.370</td>
<td>0.952</td>
</tr>
</tbody>
</table>

$\epsilon \sim \text{Logistic}(0, 1)$, $\gamma_0 = 1.0$

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1</td>
<td>0.076</td>
<td>0.312</td>
<td>0.294</td>
<td>0.920</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1</td>
<td>-0.030</td>
<td>0.837</td>
<td>0.833</td>
<td>0.932</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>0.050</td>
<td>0.536</td>
<td>0.520</td>
<td>0.922</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1</td>
<td>0.217</td>
<td>0.386</td>
<td>0.499</td>
<td>0.940</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>0.016</td>
<td>1.041</td>
<td>1.226</td>
<td>0.958</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-1</td>
<td>0.066</td>
<td>0.719</td>
<td>0.877</td>
<td>0.952</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1</td>
<td>0.019</td>
<td>0.175</td>
<td>0.185</td>
<td>0.952</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1</td>
<td>-0.042</td>
<td>0.580</td>
<td>0.564</td>
<td>0.938</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>0.001</td>
<td>0.346</td>
<td>0.344</td>
<td>0.960</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1</td>
<td>0.089</td>
<td>0.261</td>
<td>0.267</td>
<td>0.922</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>-0.062</td>
<td>0.731</td>
<td>0.774</td>
<td>0.944</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-1</td>
<td>-0.007</td>
<td>0.460</td>
<td>0.492</td>
<td>0.952</td>
</tr>
</tbody>
</table>
Table A.8: Simulation results for the parametric CRAFT model based on Bayesian estimation with 500 Monte Carlo simulations when \( n = 100 \) or 200, \( \epsilon \sim N(0, 1) \) or logistic(0, 1), \( \gamma_0 = 0.5 \) in the data generation, and we set the \( \gamma_0 \) prior to be \( \gamma_0 \sim N(0, 10) \).

\[ \epsilon \sim N(0, 1), \gamma_0 = 0.5 \]

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 100 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.031</td>
<td>0.166</td>
<td>0.163</td>
<td>0.922</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>-0.040</td>
<td>0.526</td>
<td>0.517</td>
<td>0.940</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.011</td>
<td>0.289</td>
<td>0.310</td>
<td>0.960</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.059</td>
<td>0.257</td>
<td>0.258</td>
<td>0.936</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>0.053</td>
<td>0.836</td>
<td>0.859</td>
<td>0.954</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.024</td>
<td>0.536</td>
<td>0.505</td>
<td>0.924</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.012</td>
<td>0.107</td>
<td>0.106</td>
<td>0.950</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>-0.007</td>
<td>0.331</td>
<td>0.348</td>
<td>0.964</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.010</td>
<td>0.212</td>
<td>0.206</td>
<td>0.958</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.018</td>
<td>0.175</td>
<td>0.167</td>
<td>0.946</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.060</td>
<td>0.545</td>
<td>0.577</td>
<td>0.966</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>-0.018</td>
<td>0.349</td>
<td>0.333</td>
<td>0.942</td>
</tr>
</tbody>
</table>

\[ \epsilon \sim \text{Logistic}(0, 1), \gamma_0 = 0.5 \]

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 100 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.106</td>
<td>0.367</td>
<td>0.369</td>
<td>0.946</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>-0.072</td>
<td>0.869</td>
<td>0.912</td>
<td>0.954</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.087</td>
<td>0.582</td>
<td>0.594</td>
<td>0.924</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.143</td>
<td>0.341</td>
<td>0.429</td>
<td>0.940</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.198</td>
<td>0.751</td>
<td>0.947</td>
<td>0.978</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>0.141</td>
<td>0.533</td>
<td>0.674</td>
<td>0.974</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.0</td>
<td>0.002</td>
<td>0.203</td>
<td>0.209</td>
<td>0.948</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>-0.024</td>
<td>0.621</td>
<td>0.618</td>
<td>0.940</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>-0.020</td>
<td>0.375</td>
<td>0.382</td>
<td>0.952</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.041</td>
<td>0.197</td>
<td>0.215</td>
<td>0.942</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.084</td>
<td>0.591</td>
<td>0.637</td>
<td>0.964</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>0.006</td>
<td>0.400</td>
<td>0.394</td>
<td>0.936</td>
</tr>
</tbody>
</table>
Appendix B

Supplement to Chapter 2

B.1 Explanation of the Difference Operator Matrix

We provide explanation of the difference operator matrix for the $m = 1$ and $m = 2$ cases.

First, for when $m = 1$, the penalty term includes:

$$
\sum_{l=2}^{L_n} \{\Delta^1 \alpha_l\}^2 = \sum_{l=2}^{L_n} (\alpha_l - \alpha_{l-1})^2 = (D_1 \alpha)'(D_1 \alpha) = \alpha' D_1' D_1 \alpha
$$

where

$$
D_1 \alpha = \begin{pmatrix}
-1 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \ldots & 1 & 1
\end{pmatrix}_{(L_n-1) \times L_n}
$$

Secondly, when $m = 2$, the penalty term includes

$$
\Delta^2 \alpha_l = (\alpha_l - \alpha_{l-1}) - (\alpha_{l-1} - \alpha_{l-2}) = \alpha_l - 2\alpha_{l-1} + \alpha_{l-2} \quad l = 3, \ldots, L_n
$$

Hence,

$$
\sum_{l=3}^{L_n} \{\Delta^2 \alpha_l\}^2 = (D_2 \alpha)'(D_2 \alpha) = \alpha' D_2' D_2 \alpha
$$

where
\[ D_2 \alpha = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_{L_n} \end{pmatrix} \] (B.1)

\[ = \begin{pmatrix} \alpha_3 - 2\alpha_2 + \alpha_1 \\ \alpha_4 - 2\alpha_3 + \alpha_2 \\ \alpha_5 - 2\alpha_4 + \alpha_3 \\ \vdots \\ \alpha_{L_n} - 2\alpha_{L_n-1} + \alpha_{L_n-2} \end{pmatrix} \] (B.2)

### B.2 Proof of Posterior Consistency

**Lemma 1.** The model given in Equation (2.11) is identifiable.

**Proof.** Let \( h_{\beta_1,\gamma_{1},\sigma_{1},H_{1}}(x,\delta,z) = h_{\beta_2,\gamma_{2},\sigma_{2},H_{2}}(x,\delta,z) \) for all \((x,\delta,z)\), as defined in Equation (2.11). We will work with each component of \( h_{\beta,\gamma,\sigma,H}(x,\delta,z) \) separately. Let \( \sigma_{ij}^2 = \sigma_{e_j}^2/(k_0 + 1) \), \( j = 1,2 \), so we may reparameterize the mixing distribution in terms of \( \sigma_{ij}^2 \).

First, let \( \delta = 0 \) so we work with the second component. Then for all \((x,z)\),

\[
p(\gamma_1, z) \int \Phi \left[ \frac{\log x - \beta'_1 z - \mu}{\sigma_1} \right] dH_1(\mu) = p(\gamma_2, z) \int \Phi \left[ \frac{\log x - \beta'_2 z - \mu}{\sigma_2} \right] dH_2(\mu).
\]

Now, letting \( x \to \infty \) and applying Monotone Convergence Theorem (MCT), the above equation reduces to

\[
p(\gamma_1, z) \int dH_1(\mu) = p(\gamma_1, z) \int dH_2(\mu).
\]

But \( \int_0^\infty dH(\mu) = 1 \) for any mixing distribution. Hence, \( p(\gamma_1, z) = p(\gamma_2, z) \) implies that \( \gamma_1 = \gamma_2 \) assuming that \( F_\delta(\cdot) \) is a strictly increasing function.

Next, let \( \delta = 1 \) so we work with the second component. Because \( \gamma_1 = \gamma_2 \), for all \((x,z)\), we have

\[
\int \frac{1}{\sigma_1 x} \phi \left[ \frac{\log x - \beta'_1 z - \mu}{\sigma_1} \right] dH_1(\mu) = \int \frac{1}{\sigma_2 x} \phi \left[ \frac{\log x - \beta'_2 z - \mu}{\sigma_2} \right] dH_2(\mu).
\] (B.3)
Setting \( z = 0 \), the above equation reduces to

\[
\int \frac{1}{\sigma_1 x} \phi \left[ \frac{\log x - \mu}{\sigma_1} \right] dH_1(\mu) = \int \frac{1}{\sigma_2 x} \phi \left[ \frac{\log x - \mu}{\sigma_2} \right] dH_2(\mu)
\]

(B.4)

for all \( x \in (0, \infty) \). Multiplying both sides of B.4 by \( e^{iwx} \) and integrating with respect to \( x \) we obtain

\[
\int_0^\infty \int e^{iwx} \frac{1}{\sigma_1 x} \phi \left[ \frac{\log x - \mu}{\sigma_1} \right] dH_1(\mu) dx = \int_0^\infty \int e^{iwx} \frac{1}{\sigma_2 x} \phi \left[ \frac{\log x - \mu}{\sigma_2} \right] dH_2(\mu) dx,
\]

where \( i = \sqrt{-1} \) is the imaginary unit. By change of variables, \( u_1 = (\log x - \mu)/\sigma_1 \) and \( u_2 = (\log x - \mu)/\sigma_2 \) and using Fubini’s Theorem we have

\[
\int \int e^{iw(\mu + \sigma_1 u_1)} \phi(u_1) du_1 dH_1(\mu) = \int \int e^{iw(\mu + \sigma_2 u_2)} \phi(u_2) du_2 dH_2(\mu).
\]

After re-arranging we have

\[
\int e^{iw\mu} \int e^{i\sigma_1 u_1} \phi(u_1) du_1 dH_1(\mu) = \int e^{iw\mu} \int e^{i\sigma_2 u_2} \phi(u_2) du_2 dH_2(\mu).
\]

The inside integrals are the characteristic functions for the standard normal distributions and therefore we have,

\[
\int e^{iw\mu} e^{-\frac{1}{2} \sigma_1^2 w^2} dH_1(\mu) = \int e^{iw\mu} e^{-\frac{1}{2} \sigma_2^2 w^2} dH_2(\mu).
\]

Let \( \Psi_H(w) = \int e^{iw\mu} dH(\mu) \) be the characteristic function for \( H(\cdot) \). Then the above equation simplifies to

\[
e^{-\frac{1}{2} \sigma_1^2 w^2} \Psi_{H_1}(w) = e^{-\frac{1}{2} \sigma_2^2 w^2} \Psi_{H_2}(w).
\]

Taking the logarithm on both sides of the equation we have

\[
\sigma_1^2 w^2 - 2 \log \Psi_{H_1}(w) = \sigma_2^2 w^2 - 2 \log \Psi_{H_2}(w), \tag{B.5}
\]

for all \( w \in \mathbb{R} \). Differentiating both sides of the above equation with respect to \( w \) twice results in

\[
2\sigma_1^2 - 2 \frac{d^2}{dw^2} \log \Psi_{H_1}(w) = 2\sigma_2^2 - 2 \frac{d^2}{dw^2} \log \Psi_{H_2}(w).
\]

Note that

\[
\left. \frac{d^2}{dw^2} \log \Psi_H(w) \right|_{w=0} = \left. \frac{\Psi_H''(w)\Psi_H(w) - \Psi_H'(w)^2}{\Psi_H^2(w)} \right|_{w=0}
\]

\[
= \left. \Psi_H''(0) - \Psi_H'(0)^2 \right|_{w=0}
\]

\[
= -\text{Var}_H[\mu],
\]
since $\Psi_H(0) = 1$, $\Psi_H'(0) = -iE[\mu]$, and $\Psi_H''(0) = -E[\mu^2]$. Hence we have

$$\sigma_1^2 + \text{Var}_H[\mu] = \sigma_2^2 + \text{Var}_H[\mu].$$

Recall that $\text{Var}_H[\mu] = \frac{k_0}{k_0+1}\sigma_\epsilon^2$ and hence it follows that $\sigma_1^2 = \sigma_2^2$, which implies that $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$. By Equation (B.5) this also means that $\Psi_{H_1}(w) = \Psi_{H_2}(w)$ for all $w \in \mathbb{R}$. By uniqueness of characteristic functions, $H_1(\cdot) = H_2(\cdot)$. Returning to Equation (B.3) and using DCT, we obtain for all $z$, $\beta_1^* z = \beta_2^* z$ and hence, $\beta_1 = \beta_2$. This completes the proof of identifiability.

**Lemma 2.** Consider the product topology on $(\beta, \gamma, \sigma_\epsilon, H)$, where $\beta$, $\gamma$, and $\sigma_\epsilon$, are given the usual Euclidean topology and $H$ the weak topology. On densities $h_{\beta, \gamma, \sigma_\epsilon, H}(x, \delta, z)$, put the total variation (or the $L_1$) distance defined as

$$||h_{\beta_1, \gamma_1, \sigma_\epsilon_1, H_1(\cdot)}(x, \delta, z) - h_{\beta_2, \gamma_2, \sigma_\epsilon_2, H_2(\cdot)}(x, \delta, z)|| = \int \int_0^\infty |h_{\beta_1, \gamma_1, \sigma_\epsilon_1, H_1}(x, \delta, z) - h_{\beta_2, \gamma_2, \sigma_\epsilon_2, H_2}(x, \delta, z)| \, dx \, dz.$$  

(B.6)

Then

$$||h_{\beta_1, \gamma_1, \sigma_\epsilon_1, H_1} - h_{\beta_2, \gamma_2, \sigma_\epsilon_2, H_2}|| \to 0$$  

(B.7)

if and only if $(\beta_1, \gamma_1, \sigma_\epsilon_1, H_1) \to (\beta_2, \gamma_2, \sigma_\epsilon_2, H_2)$. In other words, the variation topology on the densities is equivalent to the product topology on the indexing parameters.

**Proof.** The proof follows similarly as Lemma 2 in Ghosh and Ghosal (2006). In fact, the 'only if' portion follows directly as in Ghosh and Ghosal (2006) and using the identifiability property verified in Lemma 1. We provide details to the 'if' part, as the cure fraction and normal mixture provide a few differences in proof.

It suffices to show that the densities converge pointwise and then apply Scheffe’s theorem. Fix $(x, \delta, z)$. We show the proof for $\delta = 1$.

Because $\mu$ has a compact range, the integrand $\phi \left( \frac{\log x - \beta' z - \mu}{\sigma_\epsilon/(k_0+1)} \right) \frac{\sqrt{k_0+1}}{\sigma_\epsilon}$ of $h_{\beta, \gamma, \sigma_\epsilon, H}$ as a family of functions of $(\beta, \gamma, \sigma_\epsilon)$ indexed by $\mu$ is equicontinuous. Also $p(\gamma, \beta)$ is fixed and does not depend on $\mu$. For a given $\varepsilon > 0$, find $\nu$ large enough so that the integrands are $\varepsilon$-close for all $\mu$. For such a $\nu$,

$$|h_{\beta, \gamma, \sigma_\epsilon, H_1} - h_{\beta, \gamma, \sigma_\epsilon, H_2}| \leq |h_{\beta, \gamma, \sigma_\epsilon, H_1} - h_{\beta, \gamma, \sigma_\epsilon, H_\nu}| + |h_{\beta, \gamma, \sigma_\epsilon, H_\nu} - h_{\beta, \gamma, \sigma_\epsilon, H_2}|.$$  

(B.8)
For $n \in \mathcal{N}$, let $\{p_n\}$ and $\{A_n\}$ be sequences of real numbers. Then the following inequality holds:

$$
|p_nA_n - pA| \leq |p_n - p||A| + |p_n||A_n - A|
$$

if $|p_n| \leq 1$ (B.9)

Let $p_n = p(\gamma_n, z)$ and $A_n$ represent the density function of the error distribution, which is the same density as Ghosh and Ghosal (2006). Then $|h_{\beta, \gamma, \sigma, \epsilon, H} - h_{\beta, \gamma, \sigma, \epsilon, H_0}|$ is equivalent to $|p_nA_n - pA|$. Because $p_n(\gamma, z)$ converges pointwise to $p(\gamma, z)$, there exists an $n > n_1$ for which $|p_n - p| < \frac{\epsilon}{2|A|}$ in the first term on the RHS of Equation (B.9). Since $|p(\gamma_n, z)| \leq 1$, and as shown in the proof of Lemma 2 in Ghosh and Ghosal (2006), there also exists an $n > n_2$ for which $|A_n - A| < \frac{\epsilon}{2}$ in the second term on the RHS of Equation (B.9). Therefore the RHS is less than $\frac{\epsilon}{2|A|}|A| + \frac{\epsilon}{2} = \epsilon$. Hence, $|h_{\beta, \gamma, \sigma, \epsilon, H} - h_{\beta, \gamma, \sigma, \epsilon, H_0}| < \epsilon$. Applying Scheffe’s theorem proves the 'if' part of the theorem.

Lemma 3. For all $\varepsilon > 0$,

$$
\Pi\left\{(\beta, \gamma, \sigma, \epsilon, H) : \int \int h_{\beta, \gamma, \sigma, \epsilon, H} \log \frac{h_{\beta, \gamma, \sigma, \epsilon, H}}{h_{\beta_0, \gamma_0, \sigma_0, H_0}} \, dx \, dz < \varepsilon \right\} > 0.
$$

Proof. The log likelihood ratio is given by

$$
\Delta(\beta, \gamma, \sigma, \epsilon, H) = \begin{cases} 
\log \frac{f_{\beta, \gamma, \sigma, \epsilon, H}(x, \delta, z)}{f_{\beta_0, \gamma_0, \sigma_0, H_0}(x, \delta, z)}, & \text{if } \delta = 0, \\
\log \frac{G_{\beta, \gamma, \sigma, \epsilon, H}(x, \delta, z)}{G_{\beta_0, \gamma_0, \sigma_0, H_0}(x, \delta, z)}, & \text{if } \delta = 1.
\end{cases}
$$

We shall prove the $\delta = 1$ case. The $\delta = 0$ case is simpler. Notice that $\beta, \gamma, \sigma, \epsilon, z$ are all bounded. Thus the integrand within $f_{\beta, \gamma, \sigma, \epsilon, H}$ is bounded above and below by functions of the form $k_1 e^{c_1 x} e^{-x^2/c_2}$. Also, $p(\gamma, z)$ is bounded between 0 and 1. Taking the ratio and then logarithm, $\Delta$ in the tails (in $x$) is bounded by a multiple of a power of $x$. The rest of the proof follows as that in Lemma 3 of Ghosh and Ghosal (2006) and Theorem 3 of Ghosal et al. (1999a).

Proof of Theorem 1. Lemma 2 shows that it suffices to consider neighborhoods with respect to the $L_1$-distance. Also the space is compact. The condition of prior positivity has been verified in Lemma 3. Hence, the proof is complete.
B.3 WinBUGS Code for the Semiparametric CRAFT Model Under Right Censoring

```plaintext
for (i in 1:n.obs){
    cen[i] ~ dbern(pi[i])
    logit(pi[i]) <- inprod(gamma[1:p],x[i,1:p])
    logtime[i] ~ dnorm(logmean[i],omega0)
    logmean[i] <- eta[latent[i]]+mu[i]
    latent[i] ~ dcat(prob[]) 
    mu[i]<-inprod(beta[1:p-1],x[i,2:p])}
for (i in (n.obs+1):n){
    cen[i] ~ dbern(pistar[i])
    logit(pi[i]) <- inprod(gamma[1:p],x[i,1:p])
    mu[i]<-inprod(beta[1:p-1],x[i,2:p])
    pistar[i] <- pi[i]*max(inprod(prob[],phistar[i,]),
               step(logtime[i]-maxlogtime))
    for(l in 1:N){
           phistar[i,l]<-phi((logtime[i]-eta[l]-mu[i])*sqrt(omega0)))}

inversevariance<-1/3
inversevariancesmall<-2/3
gamma[1] ~ dnorm(0,inversevariancesmall)
for (j in 2:p){
    gamma[j] ~ dnorm(0,inversevariance)}
for (j in 1:p-1){
    beta[j] ~ dnorm(0,inversevariance)}
omega0 ~ dgamma(.1,.1)
sigma0 <- 1/sqrt(omega0)

for (k in 1:(N-1)){
    expalpha[k]<-exp(alpha[k])
    prob[k] <-exp(alpha[k])/(1+sum(expalpha[]))}
```
prob[N] <- 1 - sum(prob[1:(N-1)])
for (k in 1:(N-1-m)){
    delta[k] ~ dnorm(0, lambda1)}
for (k in 1:(N-1)){
    alpha[k] <- inprod(Dcoef[k,], delta[])}
### B.4 Simulation Results

Table B.1: Simulation results for the semiparametric CRAFT model based on Bayesian estimation with 1000 Monte Carlo simulations when \( n = 100 \) or 200, \( \gamma_0 = 0.5 \) or 1.0 in the data generation, and \( \epsilon \) is generated from a symmetric logistic distribution.

\[
\epsilon \sim \text{Logistic, } \gamma_0 = 0.5
\]

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \gamma_0 )</th>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.14</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
<td>0.14</td>
<td>0.22</td>
<td>0.68</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>0.14</td>
<td>0.23</td>
<td>0.73</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>200</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.09</td>
<td>0.16</td>
<td>0.50</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.10</td>
<td>0.16</td>
<td>0.53</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.96</td>
<td>0.94</td>
<td>0.96</td>
<td>0.95</td>
</tr>
</tbody>
</table>

\[
\epsilon \sim \text{Logistic, } \gamma_0 = 1.0
\]

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \gamma_0 )</th>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.19</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.12</td>
<td>0.25</td>
<td>0.70</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.13</td>
<td>0.25</td>
<td>0.77</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>200</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.09</td>
<td>0.17</td>
<td>0.54</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.09</td>
<td>0.17</td>
<td>0.57</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Table B.2: Simulation results for the semiparametric CRAFT model based on Bayesian estimation with 1000 Monte Carlo simulations when \( n = 100 \) or 200, \( \gamma_0 = 0.5 \) or 1.0 in the data generation, and \( \epsilon \) is generated from a skewed extreme value distribution.

\( \epsilon \sim \text{Extreme Value}, \gamma_0 = 0.5 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>0.01</td>
<td>0.22</td>
<td>0.24</td>
<td>0.96</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.00</td>
<td>0.13</td>
<td>0.14</td>
<td>0.96</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>0.5</td>
<td>0.02</td>
<td>0.22</td>
<td>0.23</td>
<td>0.96</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.13</td>
<td>0.68</td>
<td>0.73</td>
<td>0.96</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>0.00</td>
<td>0.44</td>
<td>0.45</td>
<td>0.95</td>
</tr>
</tbody>
</table>

| 200   |       |      |      |     |     |
| \( \beta_1 \) | -1.0 | 0.01 | 0.16 | 0.16 | 0.95 |
| \( \beta_2 \) | 1.0  | 0.00 | 0.09 | 0.09 | 0.96 |
| \( \gamma_0 \) | 0.5  | 0.00 | 0.16 | 0.16 | 0.94 |
| \( \gamma_1 \) | 1.0  | -0.04 | 0.50 | 0.53 | 0.96 |
| \( \gamma_2 \) | -1.0 | 0.01 | 0.31 | 0.32 | 0.95 |

\( \epsilon \sim \text{Extreme Value}, \gamma_0 = 1.0 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-1.0</td>
<td>0.01</td>
<td>0.20</td>
<td>0.22</td>
<td>0.97</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.0</td>
<td>0.00</td>
<td>0.12</td>
<td>0.13</td>
<td>0.97</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>1.0</td>
<td>0.02</td>
<td>0.25</td>
<td>0.25</td>
<td>0.95</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>1.0</td>
<td>-0.18</td>
<td>0.70</td>
<td>0.77</td>
<td>0.97</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-1.0</td>
<td>0.02</td>
<td>0.46</td>
<td>0.48</td>
<td>0.96</td>
</tr>
</tbody>
</table>

| 200   |       |      |      |     |     |
| \( \beta_1 \) | -1.0 | 0.00 | 0.14 | 0.15 | 0.95 |
| \( \beta_2 \) | 1.0  | 0.00 | 0.08 | 0.09 | 0.97 |
| \( \gamma_0 \) | 1.0  | 0.02 | 0.17 | 0.17 | 0.96 |
| \( \gamma_1 \) | 1.0  | -0.08 | 0.54 | 0.57 | 0.96 |
| \( \gamma_2 \) | -1.0 | 0.03 | 0.34 | 0.34 | 0.94 |
Table B.3: Simulation results for the semiparametric CRAFT model based on Bayesian estimation with 1000 Monte Carlo simulations when $n = 100$ or 200, $\gamma_0 = 0.5$ or 1.0 in the data generation, and $\epsilon$ is generated from a bimodal distribution.

$\epsilon \sim \text{Bimodal, } \gamma_0 = 0.5$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.16</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.11</td>
<td>0.23</td>
<td>0.66</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.13</td>
<td>0.23</td>
<td>0.73</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>0.98</td>
<td>0.94</td>
<td>0.97</td>
<td>0.95</td>
</tr>
</tbody>
</table>

$\epsilon \sim \text{Bimodal, } \gamma_0 = 1.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.12</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.10</td>
<td>0.24</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.11</td>
<td>0.25</td>
<td>0.77</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>-1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.10</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.18</td>
<td>0.54</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>
B.5 Figures from the Breast Cancer Data Set

Figure B.1: Breast cancer data: posterior means of the mixing weights using the semiparametric CRAFT model
Figure B.2: Breast cancer data: trace and posterior density curves for both the latency and incidence statistics associated with treatment A or B, clinical stage indication, and number of lymph nodes using the semiparametric CRAFT model.
Appendix C

Supplement to Chapter 3

C.1 Proof of Posterior Consistency

The following three lemmas are used in the proof of Theorem 2.

**Lemma 4.** Let $f_c$ be a density function and let $S_0$ and $F_0$ be survival and cumulative distribution functions. Consider $S(\cdot)$ as defined in Equation (3.3) and $F(\cdot) = 1 - S(\cdot)$ where

1. $i \int |x - \mu| f_c(x) dH(\mu) < \infty$
2. $ii \int f_c(x) |\log S_0(x)| dx < \infty$
3. $iii \int f_c(x) |\log F_0(x)| dx < \infty$

Then

1. $\lim_{\gamma \to 0} \int f_c(x) \log \frac{S_0(x)}{S(x - \gamma)} dx = \int f_c(x) \log \frac{S_0(x)}{S(x)} dx$
2. $\lim_{\gamma \to 0} \int f_c(x) \log \frac{F_0(x)}{F(x - \gamma)} dx = \int f_c(x) \log \frac{F_0(x)}{F(x)} dx$.

**Proof.** The general line of proof for each of the lemma’s two conclusions is to (a) verify the conditions of the dominated convergence theorem (DCT) in order to bring the limit inside the integral and (b) show that the limit of the integrand converges to the correct quantity.

For $1.$, we need to show $\lim_{\gamma \to 0} \int f_c(x) \log \frac{S_0(x)}{S(x - \gamma)} dx = \int f_c(x) \log \frac{S_0(x)}{S(x)} dx$. To apply the DCT, we need to show that $|\log S(x - \gamma)| \leq M(x, \mu)$ for all $\gamma \leq \gamma_0$, $\gamma_0 > 0$, and any $\mu$ such that $\int M(x, \mu) f_c(x) dx < \infty$. 
Because \( S(x) \leq 1 \) for all \( x \in \mathbb{R} \), using Jensen’s inequality for the convex function \(-\log(\cdot)\), because \( x + \log(1 + e^{-x}) \) is an increasing function of \( x \), and since \( \log(1 + e^{-x}) \leq |x| + \log(2) \) for all \( x \in \mathbb{R} \), it follows that

\[
| \log S(x - \gamma) | = -\log S(x - \gamma) \\
= -\log \int \frac{e^{-(x-\mu-\gamma)/\lambda}}{1 + e^{-(x-\mu-\gamma)/\lambda}} dH(\mu) \\
= -\log E_\mu \left\{ \frac{e^{-(x-\mu-\gamma)/\lambda}}{1 + e^{-(x-\mu-\gamma)/\lambda}} \right\} \\
\leq E \left\{ -\log \frac{e^{-(x-\mu-\gamma)/\lambda}}{1 + e^{-(x-\mu-\gamma)/\lambda}} \right\} \\
= \int \left\{ (x - \mu - \gamma)/\lambda + \log \{1 + e^{-(x-\mu-\gamma)/\lambda}\} \right\} dH(\mu) \\
\leq \int \left\{ (|x - \mu| + \gamma_0)/\lambda + \log \{1 + e^{-(|x - \mu| + \gamma_0)/\lambda}\} \right\} dH(\mu) \\
\leq \int \left\{ (|x - \mu| + \gamma_0)/\lambda + (|x - \mu| + \gamma_0)/\lambda + \log(2) \right\} dH(\mu) \\
= M(x, \mu).
\]

Then,

\[
\int M(x, \mu) f_c(x) dx = \int \int \left\{ (|x - \mu| + \gamma_0)/\lambda + (|x - \mu| + \gamma_0)/\lambda + \log(2) \right\} dH(\mu) f_c(x) dx < \infty
\]

where the last line holds because of condition \( i \). Hence, based on conditions \( i \) and \( ii \), we may apply the DCT as follows;

\[
\lim_{\gamma \to 0} \int f_c(x) \log \left\{ \frac{S_0(x)}{S(x - \gamma)} \right\} dx = \lim_{\gamma \to 0} \int \left\{ f_c(x) \log S_0(x) - f_c(x) \log S(x - \gamma) \right\} dx \\
= \int f_c(x) \log S_0(x) dx - \int f_c(x) \lim_{\gamma \to 0} \log S(x - \gamma) dx \\
= \int f_c(x) \log \frac{S_0(x)}{S(x)} dx.
\]

Because \( F_0 \) may be written in a similar format as \( S_0 \), the proof follows similarly for \( 2. \) using conditions \( i \) and \( iii \) and applying the DCT.
Lemma 5. For the AFT model under the conditions of Theorem 2, there is an exponentially consistent sequence of tests $H_0 : (F, \beta) = (F_0, \beta_0)$ against $H_1 : F \notin U_{F_0}$, where $U_{F_0}$ is a weak neighborhood of $F_0(\cdot)$.

Lemma 6. For the AFT model under the conditions of Theorem 2, there is an exponentially consistent sequence of tests $H_0 : (F, \beta) = (F_0, \beta_0)$ against $H_1 : \sup_{x \in \mathbb{R}} |F(x) - F_0(x)| < \Delta, \|\beta - \beta_0\| > \delta$, where $U_{F_0}$ is a weak neighborhood of $F_0(\cdot)$.

Proof. Proofs for Lemmas 5 and 6 can be found in Wu (2009) where lemmas 5 and 6 correspond to Lemmas 17 and 18.

Proof of Theorem 2. Appealing to Schwartz (1965)’s celebrated theorem, posterior consistency can be established by verifying two sufficient conditions. Namely,

1. the existence of a strictly unbiased test for $H_0 : (F, \beta) = (F_0, \beta_0)$ versus its complement $H_1 : F \notin U_{F_0}$ or $\|\beta - \beta_0\| > \delta$ where $U_{F_0}$ is a weak neighborhood of $F_0(\cdot)$,

2. the Kullback-Leibler support condition on $\Pi$ at $(f_0, \beta_0)$ holds for this AFT model with mixture of logistic distributions

Wu (2009)’s Theorem 1 clearly states these two conditions while Amewou-Atisso et al. (2003)’s Theorem 2.1 extends the result to independent, non-identically distributed response variables and non-random covariates. Notice that Lemma 5 and Lemma 6 provide an exponentially consistent test for $H_0 : (F, \beta) = (F_0, \beta_0)$ versus $H_1 : F \notin U_{F_0}$ or $\|\beta - \beta_0\| > \delta$ since $P\{(F \notin U_{F_0}) \cup (\|\beta - \beta_0\| > \delta)\} = P(F \notin U_{F_0}) + P(\|\beta - \beta_0\| > \delta \& F \in U_{F_0})$. Hence, the testing condition is verified. Also note, power properties for the constructed tests that are stated in Wu (2009)’s Theorem 1 hold here for the entire parameter space. Therefore, it remains to verify that the Kullback-Leibler property holds. That is, we must show,

$$\Pi\left\{ \int \int F_0(x - \beta_0^T z) f_c(x) \log \frac{F_0(x - \beta_0^T z)}{F(x - \beta_0^T z)} dxdQ(z) + \int \int S_0(x - \beta_0^T z) f_c(x) \log \frac{S_0(x - \beta_0^T z)}{S(x - \beta_0^T z)} dxdQ(z) < \epsilon \right\} > 0 \quad (C.1)$$

for any $\epsilon > 0$. We begin with the second piece of Inequality (C.1). We first show that

$$\Pi_\lambda \times \Pi_H \left\{ \int f_c(x + \xi) \log \frac{S_0(x)}{S(x)} dx < \epsilon \right\} > 0, \quad (C.2)$$
holds for any $\xi \in \mathbb{R}$.

Define

$$f_m(x) = \begin{cases} t_m f_0(x), & -m < x < m, \\ 0, & \text{otherwise,} \end{cases}$$

where $m \geq 1$, $t_m^{-1}(x) = \int_{|x|<m} f_0(x)dx$, $\lambda_m = m^{-\eta}$ for some $\eta > 0$, $F_m$ is the probability measure corresponding to $f_m$, and $H_m = F_m \times \delta(\lambda_m)$, where $\delta(\cdot)$ is the degenerate distribution. Since $H_m$ is compactly supported and using the transformation $\mu = x - \lambda_mu$,

$$f_{H_m}(x) = \int \frac{e^{-(x-\mu)/\lambda_m}}{(1 + e^{-(x-\mu)/\lambda_m})^2} dF_m(\mu)$$

$$= t_m \int_{|\mu|<m} \frac{e^{-(x-\mu)/\lambda_m}}{(1 + e^{-(x-\mu)/\lambda_m})^2} f_0(\mu) d\mu$$

$$= \int_{|x-\lambda_mu|<m} \frac{e^{-u}}{(1 + e^{-u})^2} f_0(x - \lambda_mu) du.$$

Since $0 < f_0(x) < M$ for all $x$ by condition 2, the above integrand is bounded by $\frac{e^{-u}}{(1 + e^{-u})^2} M$. Also $\int \frac{e^{-u}}{(1 + e^{-u})^2} M du < \infty$ since $\frac{e^{-u}}{(1 + e^{-u})^2}$ is the proper logistic density function and $f_0(x - \lambda_mu) \to f_0(x)$ as $\lambda_m \to 0$. Therefore, we may apply the DCT and obtain $f_{H_m}(x) \to f_0(x)$ by Scheffe’s theorem, $S_{H_m} \to S_0$ as $\lambda_m \to 0$ as well, where

$$S_{H_m}(x) = \int_0^x f_{H_m}(s) ds$$

$$= \int_0^x \int \frac{e^{-(s-\mu)/\lambda_m}}{(1 + e^{-(s-\mu)/\lambda_m})^2} f_0(\mu) d\mu ds$$

$$= \int \int_0^x \frac{e^{-(s-\mu)/\lambda_m}}{(1 + e^{-(s-\mu)/\lambda_m})^2} ds f_0(\mu) d\mu$$

$$= \int \frac{e^{-(x-\mu)/\lambda_m}}{1 + e^{-(x-\mu)/\lambda_m}} f_0(\mu) d\mu$$

$$= t_m \int_{|x|<m} \frac{e^{-(x-\mu)/\lambda_m}}{1 + e^{-(x-\mu)/\lambda_m}} f_0(\mu) d\mu.$$

In order to prove Inequality (C.2), we first will show

$$\int_{-\infty}^{\infty} f_c(x + \xi) \log \frac{S_0(x)}{S_{H_m}(x)} dx \to 0$$

as $m \to \infty$ for any $\xi \in \mathbb{R}$. Since $S_0$ is a proper survival function, $S_0(x) \to 1$ as $x \to -\infty$. Also recall $S_{H_m}(x) \to S_0(x)$ as $\lambda_m \to 0$. Therefore, $\log \frac{S_0(x)}{S_{H_m}(x)} \to 0$ as $x \to -\infty$ and $\lambda_m \to 0$. 
Then

$$\left| \log \frac{S_0(x)}{S_{H_m}(x)} \right| \leq \begin{cases} 
\epsilon, & x \in (-\infty, -M_0], \\
M, & x \in [-M_0, 0],
\end{cases}$$

for some $M_0 > 0$. Therefore we may apply the DCT to obtain $\int_{-\infty}^{0} f_c(x+\xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx \to 0$ as $\lambda_m \to 0$. Hence, it remains to show $\int_{0}^{\infty} f_c(x+\xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx \to 0$.

Notice,

$$\int_{0}^{\infty} f_c(x+\xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx = \int_{0}^{m} f_c(x+\xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx + \int_{m}^{\infty} f_c(x+\xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx.$$

Let $x > m$. Because $e^{-\lambda_m(x-\mu)} / \{1+e^{-\lambda_m(x-\mu)}\}$ is a monotone increasing function of $\mu$,

$$S_{H_m}(x) = t_m \int_{|\mu|<m} \frac{e^{-(x-\mu)/\lambda_m}}{1+e^{-(x+\mu)/\lambda_m}} f_0(\mu) \, d\mu \geq t_m \int_{|\mu|<m} \frac{e^{-(x+\mu)/\lambda_m}}{1+e^{-(x+\mu)/\lambda_m}} f_0(\mu) \, d\mu = t_m \frac{e^{-(x+\mu)/\lambda_m}}{1+e^{-(x+\mu)/\lambda_m}} \int_{|\mu|<m} f_0(\mu) \, d\mu = \frac{e^{-(x+\mu)/\lambda_m}}{1+e^{-(x+\mu)/\lambda_m}} \int_{|\mu|<m} f_0(\mu) \, d\mu \geq \frac{e^{-2x^{\eta+1}}}{1+e^{-2x^{\eta+1}}}. \quad (C.4)$$

The last inequality holds because $e^{-s} / (1 + e^{-s})$ is a monotone decreasing function of $s$ and $(x+m)m^{\eta} < 2x^{\eta+1}$ when $x > m$ and $m^{\eta} > 0$.

Next, let $x < m$ and fix $\delta > 0$, where $\phi_m(x) = \inf_{|x-s|<\delta} f_0(s)$. Then since
$t_m \geq 1$, $\phi_m(x) \geq \ldots \geq \phi_1(x)$, and the logistic survival function is a non-negative function,

$$S_{H_m}(x) = t_m \int_{|\mu|<m} \frac{e^{-(x-\mu)/\lambda_m}}{1 + e^{-(x-\mu)/\lambda_m}} f_0(\mu) d\mu$$

$$\geq t_m \int_{|\mu|<m \cap \{|x-\mu|<\delta \lambda_m\}} \frac{e^{-(x-\mu)/\lambda_m}}{1 + e^{-(x-\mu)/\lambda_m}} f_0(\mu) d\mu$$

$$\geq t_m \phi_m(x) \int_{|\mu|<m \cap \{|x-\mu|<\delta \lambda_m\}} \frac{e^{-(x-\mu)/\lambda_m}}{1 + e^{-(x-\mu)/\lambda_m}} d\mu$$

$$\geq t_m \phi_m(x) \int_{|\mu|<m \cap \{|x-\mu|<\delta \lambda_m\}} e^{-\delta} \frac{1}{1 + e^{-\delta}} d\mu$$

$$= t_m \phi_m(x) \int_{\min(m,x+\delta \lambda_m)} \max(-m,x-\delta \lambda_m)} \frac{1}{1 + e^{-\delta}} d\mu$$

$$\geq \phi_1(x)c_1,$$  \hspace{1cm} (C.5)

where $c_1 = \frac{e^{-\delta}}{1 + e^{-\delta}} 2\delta \lambda_m$ and the last inequality holds when $\delta$ is chosen small so that $x + \delta \lambda_m \leq m$ and $x - \delta \lambda_m \geq -m$. Combining Inequalities (C.4) and (C.5),

$$S_{H_m}(x) \geq \begin{cases} 
\phi_1(x)c_1, & x < m, \\
\frac{e^{-2x\eta+1}}{1 + e^{-2x\eta+1}} & x > m.
\end{cases}$$

Also, $S_{H_m}(x) \leq 1$. Therefore,

$$\log S_0(x) \leq \log \frac{S_0(x)}{S_{H_m}(x)} \leq \xi(x) = \begin{cases} 
\log \frac{S_0(x)}{\phi_1(x)c_1}, & x < m, \\
\log \frac{S_0(x)(1 + e^{-2x\eta+1})}{e^{-2x\eta+1}} & x > m
\end{cases}$$

$$|\log \frac{S_0(x)}{S_{H_m}(x)}| \leq \max \left\{ \xi(x), |\log S_0(x)| \right\}$$  \hspace{1cm} (C.6)

Looking at the first piece of the right-hand side (RHS) of Inequality (C.6):

$$\int_{0}^{\infty} \xi(x)f_c(x+\xi)dx = \int_{0}^{m} f_c(x+\xi) \log \frac{S_0(x)}{c_1 \phi_1(x)} dx + \int_{m}^{\infty} f_c(x+\xi) \log \frac{S_0(x)(1 + e^{-2x\eta+1})}{e^{-2x\eta+1}} dx.$$  \hspace{1cm} (C.7)

The first term on the RHS of Equation (C.7) is finite from condition 6 while the second term on the RHS is finite based on conditions 3 and 7. Also, from the second piece of the RHS of Inequality (C.6), $\int |\log S_0(x)| f_c(x + \xi) dx$ is finite based on condition 3. Therefore
we may apply the dominated convergence theorem (DCT) as follows:

\[
\lim_{m \to \infty} \int_0^\infty f_c(x + \xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx = \int_0^\infty f_c(x + \xi) \lim_{m \to \infty} \log \frac{S_0(x)}{S_{H_m}(x)} \, dx \\
= \int_0^\infty f_c(x + \xi) \log \frac{S_0(x)}{S_0(x)} \, dx \\
= 0
\]

for any \( \xi \in \mathbb{R} \).

Since \( \int f_c(x + \xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx \) is a random sequence of numbers that converges to 0, there exists an \( H_m \) with large enough \( m \) for which \( \int f_c(x + \xi) \log \frac{S_0(x)}{S_{H_m}(x)} \, dx < \epsilon \) has positive probability. Hence, Inequality (C.2) holds. And lastly, by Inequality (C.2) and part 2 of Lemma 4 (valid by conditions 3, 4 and 7), the second piece of Inequality (C.1) is true, namely

\[
\Pi \left\{ \int \int S_0(x - \beta_0^T z) f_c(x) \log \left\{ \frac{S_0(x - \beta_0^T z)}{S(x - \beta_0^T z)} \right\} \, dx \, dQ(z) < \epsilon \right\} > 0. \quad (C.8)
\]

It remains to verify the first piece of Inequality (C.1), related to the Kullback-Leibler distance between \( F \) and \( F_0 \). Similar to \( S_{H_m} \), define \( F_{H_m} = t_m \int |x| < \mu \{1 + e^{-(x-\mu)/\lambda_m}\}^{-1} f_0(\mu) \, d\mu \). Using similar arguments as above, it can be verified that

\[
F_{H_m}(x) \geq \begin{cases} 
\phi_1(x)c_2, & x < m, \\
(1 + e^{2|x|^\eta})^{-1} & x > m,
\end{cases}
\]

where \( \phi_m(x) = \inf_{|x-s| < \delta \lambda_m} f_0(s) \) and \( c_2 = (1 + e^{-\delta})^{-1}2\delta \lambda_m \) for some fixed \( \delta \). Then using the same arguments as above but in terms of \( F_{H_m} \), and by using part 2 of Lemma 4,

\[
\Pi \left\{ \int \int F_0(x - \beta_0^T z) f_c(x) \log \left\{ \frac{F_0(x - \beta_0^T z)}{F(x - \beta_0^T z)} \right\} \, dx \, dQ(z) < \epsilon \right\} > 0 \quad (C.9)
\]

Therefore combining Inequalities (C.8) and (C.9), Inequality (C.1) holds.
C.2 WinBUGS Code for the Semiparametric AFT Model
Under Interval Censoring

# Type 1 Interval-Censored Data
for (i in 1:n){
    delta[i] ~ dbern(pistar[i])
    mu[i]<-inprod(beta[1:p],x[i,1:p])
    logit(pistar[i]) <- (logcen[i]-mu[i]-u[i])/lambda
    u[i]<-eta[latent[i]]
    latent[i] ~ dcat(prob[])}

for (k in 1:N){alpha[k]<-2/N}
prob[1:N] ~ ddirch(alpha[])
u0<-.65
u1<-.75
lambda<-pow(n,-1/5)*kappa
kappa ~ dunif(u0,u1)
inversevariance<-1/3
for (k in 1:p){beta[k] ~ dnorm(0,inversevariance)}

# Type 2 Interval-Censored Data
for (i in 1:n){
    delta[i, 1 :3] ~ dmulti( pistar[i, 1 : 3],1)
    mu[i]<-inprod(beta[1:p-1],x[i,2:p])
    logit(pistar[i,1]) <- (logleft[i]-mu[i]-u[i])/lambda
    logit(pistar[i,3]) <- -(logright[i]-mu[i]-u[i])/lambda
    pistar[i,2] <- 1-pistar[i,1]-pistar[i,3]
    u[i]<-eta[latent[i]]
    latent[i] ~ dcat(prob[])}

for (k in 1:N){alpha[k]<-2/N}
prob[1:N] ~ ddirch(alpha[])
u0<-.6
u1<-.8
lambda<-pow(n,-1/5)*kappa
kappa ~ dunif(u0,u1)
inversevariance<-1/3
for (k in 1:p){beta[k] ~ dnorm(0,inversevariance)}
### C.3 Simulation Results

Table C.1: Simulation results for the semiparametric AFT model under case 1 interval censoring with 100 Monte Carlo simulations. Sample sizes are $n = 200$ or 400, and $\epsilon$ is generated from normal, logistic, extreme value or bimodal distributions.

<table>
<thead>
<tr>
<th>$\epsilon \sim$ Normal</th>
<th>$n$</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.05</td>
<td>0.24</td>
<td>0.26</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.04</td>
<td>0.14</td>
<td>0.16</td>
<td>0.93</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.01</td>
<td>0.18</td>
<td>0.17</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.02</td>
<td>0.12</td>
<td>0.10</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\epsilon \sim$ Logistic</th>
<th>$n$</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.04</td>
<td>0.23</td>
<td>0.25</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.02</td>
<td>0.14</td>
<td>0.15</td>
<td>0.96</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.02</td>
<td>0.17</td>
<td>0.17</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.02</td>
<td>0.12</td>
<td>0.10</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\epsilon \sim$ Extreme Value</th>
<th>$n$</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.05</td>
<td>0.24</td>
<td>0.25</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.06</td>
<td>0.15</td>
<td>0.15</td>
<td>0.94</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.04</td>
<td>0.17</td>
<td>0.17</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.06</td>
<td>0.11</td>
<td>0.10</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\epsilon \sim$ Bimodal</th>
<th>$n$</th>
<th>TRUE</th>
<th>BIAS</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.09</td>
<td>0.21</td>
<td>0.27</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.11</td>
<td>0.18</td>
<td>0.16</td>
<td>0.84</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>-1.00</td>
<td>-0.02</td>
<td>0.17</td>
<td>0.17</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.07</td>
<td>0.10</td>
<td>0.10</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Table C.2: Simulation results for the semiparametric AFT model under case 2 interval censoring with 100 Monte Carlo simulations. Sample sizes are $n = 200$ or $400$, and $\epsilon$ is generated from normal, logistic, extreme value or bimodal distributions.

$\epsilon \sim \text{Normal}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>True</th>
<th>Bias</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>$-0.01$</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>$0.01$</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>$-0.01$</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>$-0.01$</td>
<td>0.09</td>
<td>0.08</td>
</tr>
</tbody>
</table>

$\epsilon \sim \text{Logistic}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>True</th>
<th>Bias</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>0.00</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.00</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>0.01</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>$-0.02$</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

$\epsilon \sim \text{Extreme Value}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>True</th>
<th>Bias</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>$-0.01$</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.04</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>$-0.02$</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.02</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

$\epsilon \sim \text{Bimodal}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>True</th>
<th>Bias</th>
<th>MCSE</th>
<th>ESE</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>$-0.04$</td>
<td>0.17</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td>400</td>
<td>Beta 1</td>
<td>$-1.00$</td>
<td>$-0.03$</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Beta 2</td>
<td>1.00</td>
<td>0.05</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>
C.4 Figures from the Breast Cosmesis Data Set

Figure C.1: Interval-censored breast cosmesis data: trace and posterior density curve for beta (treatment parameter) using the semiparametric AFT model
Figure C.2: Interval-censored breast cosmesis data: posterior means of the mixing weights using the semiparametric AFT model